AD_____

Award Number:
W81XWH-10-1-0299

TITLE:
Uncovering the Hidden Molecular Signatures of Breast Cancer

PRINCIPAL INVESTIGATOR:
Robert Lesurf

CONTRACTING ORGANIZATION:
McGill University, Montreal, Quebec, Canada

REPORT DATE:
May 2013

TYPE OF REPORT:
Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
May 2013	Annual Summary	1 May 2010 – 30 April 2013

4. TITLE AND SUBTITLE
Uncovering the Hidden Molecular Signatures of Breast Cancer

5a. CONTRACT NUMBER
W81XWH-10-1-0299

5b. GRANT NUMBER
W81XWH-10-1-0299

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Robert Lesurf

E-Mail: robert.lesurf@gmail.com

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
McGill University

Montreal H3A 2T5
Canada

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Breast cancer is a heterogeneous disease, consisting of at least five transcriptional subtypes described by distinct, but poorly understood, molecular profiles. As the cause, aggressiveness, and outcome vary greatly between patients, it is essential to characterize the different ways in which the disease can grow and spread. Transcriptional subtyping operates by capturing the 'loudest' molecular events within a tumor. While these events are both biologically and clinically important, they only represent a fraction of the total cellular pathways and responses. Subtle information is overshadowed by these responses. Such information lies orthogonally to the subtypes, and may be of equal or greater clinical importance. We have constructed a framework to address this challenge. Our methodology is different from what has been done in the past, because it is able to break down tumors using individual signatures. The analysis is done on a large-scale, and does not require tumors to be binned into distinct classes. In a similar way, murine and cell-line models have been analyzed, allowing us to determine which models best reflect the human disease, and in what way. This is in turn allowing us to understand how different tumor processes work together, and to refine our models to better reflect the human disease. We've aimed to produce an open and accessible framework that will be used to quickly and thoroughly understand the processes that are at play in new tumour cases. This framework will have immediate research applications through the generation of better models for breast cancer. Ultimately, we intend for it to provide patients with more accurate, appropriate, and personalized treatments.

15. SUBJECT TERMS
Breast Cancer

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	54	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Introduction

Breast cancer is a heterogeneous disease, with at least five intrinsic subtypes including the luminal A and luminal B (estrogen receptor alpha positive; ESR+), Her2+ (v-erb-b2 erythroblastic leukemia viral oncogene homolog 2 positive), basal (ESR-, Her2-), and normal-like patient groups[1-3]. These subtypes exhibit distinct differences in their molecular signaling cascades, stress responses, and in the types of cells present within the tumor. For example, the luminal subtypes of breast cancer display a strong estrogen-signaling component, while the Her2+ subtype reflects the downstream response of receptor tyrosine kinase activation. Furthermore, recent studies have suggested that there may be greater heterogeneity amongst tumor subtypes than was previously understood[4-6]. A more complete understanding of tumor pathways and responses is needed to fully determine the reasons for treatment failure and disease recurrence. To date, however, we lack a comprehensive analysis of those processes within the tumor that are associated with outcome (or other histopathological/clinical variables), and whether they are dependent or independent of the tumor subtype.

Our central hypotheses are that each tumor can be defined as a collection of molecular processes, that there exist processes that can be used to predict patient outcome regardless of subtype and other recognized clinical variables, and that there exist a disjoint set of processes that predict prognosis within each subtype. Moreover, we argue that the identity of these processes can be inferred through the combined use of our de novo bioinformatics framework entitled Breast Signature Analysis Tool (BreSAT) and our catalogue of transcriptional signatures (entitled BreSAT-DB) that have been collected from literature and resources such as GeneSigDB[7] and MSigDB[8], but carefully modified and augmented to reflect the specific biologies of the breast environment.

We have applied BreSAT and it's associated catalogue BreSAT-DB to thousands of breast tumor samples and models of the disease. This has allowed us to identify novel pathways, processes, responses, and cell types that are of interest to disease progression and outcome, in addition to the identification of highly correlated processes that share little or no biological commonalities. These processes of interest were largely recapitulated in the models investigated thus far, although we identify various elements with relevance to the human disease that are currently lacking in the models. In one specific example, we've used our framework in combination with experimental validation, to identify that synergy between the oncogene MET and loss of p53 (tumor protein p53) lead to a tumor phenotype that reflects the human claudin-low subclass of breast cancer[9]. Together, these discoveries are leading to a more comprehensive and complete view of breast cancer and the generation of more accurate disease models.

Body

Task 1. Complete course requirements (year 1):
1a. BIOC 603: Genomics and Gene Expression (year 1).

All required PhD coursework was successfully completed in year 1. Other program requirements to date, including research seminars 1 & 2 (junior seminar and PhD proposal respectively) were also successfully completed.

Task 2. Development of breast cancer-specific signatures (year 1):
2a. Acquire signatures from literature and databases (year 1).
2b. Filter collection based on relevancy (year 1).
2c. Agglomerate signatures representing high biological similarity (year 1).
2d. Refine genes according to behavior in breast-related datasets (year 1).
Milestone #1 Publication (year 1).

A major component of our framework involved the collection and formatting of molecular signatures, along with the development of an appropriate ontological annotation. We have termed this highly curated signature database Breast Signature Analysis Tool Database (BreSAT-DB). Signatures are typically a set of genes that have been determined to be differentially perturbed in response to either a specific molecular event (e.g. overexpression of ESR), or are markers of a specific cell type (e.g. macrophages versus pericytes versus endothelial cells). Signature databases such as GeneSigDB[7] and MSigDB[8] exist, and contain thousands of such signatures. However, these signatures have been generated in a variety of organisms, tissues, cell types, and with different techniques. Thus, many of these signatures may not accurately recapitulate the target biology in human clinical breast samples. Furthermore, in some cases, multiple signatures exist for what are meant to be the same biological processes. This creates challenges downstream in the analysis, as separate signatures that represent the same general process or cell type may contain a dissimilar set of genes, which exhibit different expression patterns in human breast cancer data, and ultimately lead to contradictory conclusions. For these reasons, we have refined and annotated thousands of available signatures with features such as the species and tissue they were generated in, as well as their general category (e.g. whether they are used to define a particular cell type, biological response, or a broad prognostic response). Within each of these categories, the signatures are further sub-classified as appropriate (e.g. signatures that define biological responses are sub-classified into one of ten hallmarks of cancer[10]). Our categorizations are intended to allow for the first broad attempt at comprehensively dissecting breast tumors into a set of individual cellular and mechanistic components, and may be further refined and expanded by the community over time. BreSAT-DB now contains approximately 6500 signatures, which have been formatted for direct computational analysis and individually curated according to features of interest with respect to breast cancer.

In addition, we have generated a data compendium now containing ~20,000 human patient samples related to breast cancer, along with their associated histopathological/clinical data. Our compendium has been stratified by stages of disease progression (e.g. normal tissue, DCIS, IDC, metastases, etc.), type of sample (e.g. whole tumor versus cell-specific tissue derived by laser capture microdissection), adjuvant and neoadjuvant treatments, and type of data (e.g. gene expression microarrays, aCGH, miRNA, etc.). The collection involved a rigorous process of normalization and harmonization. Clinical parameters have been carefully matched to determine, for example, whether recurrence is measured as a local or distant event that takes place in a common 5- or 10-year time frame. This ensures that clinical information is directly comparable from one dataset to the next, and allowed us to develop automated tools for analyzing the data. While our focus has been on human data, we also have a sizable compendium of models for the disease, including murine tumors and human cell lines.

The collection and annotation of our database and compendium has been relatively straightforward, albeit a time consuming process. Years 2 and 3 oversaw minor updates to the size of the database (approximately 500 new signatures added), and further refinement of all signature annotations. In addition, our data compendium has expanded to include ~10,000 additional samples, and we are continuously collecting data from other platforms, now including next generation sequencing. Outside publications involving signature collection and analysis by other groups[11-13] required that we re-evaluate, re-write, and expand aspects of our manuscript in order to differentiate ourselves and highlight the unique advantages BreSAT-DB provides for breast cancer research. This has included a detailed demonstration that signatures developed in the breast are more informative than equivalent signatures developed in other tissue types, when applied to breast cancer datasets. Furthermore, breast-derived signatures contain genes that tend to be more highly correlated with one-another, suggesting that BreSAT-DB is more accurate and approximate than general-purpose signature databases for use in breast cancer research.

To aid with the distribution of the framework to general-purpose users, year 3 oversaw construction on a website that is able to dynamically accept point-and-click commands from users. This website allows users to explore the signatures and datasets in BreSAT, and through backend integration of the website with R, users may apply and compare signatures of interest to desired subsets of the datasets. The publication originally intended for Task 2 has now been merged with the publication intended for Task 4, which will include public distribution of the framework.

Task 3. Refinement of statistical methodology (year 1):
3a. Statistic for cohesiveness of subtypes (year 1).
3b. Statistic for association with survival/recurrence (year 1).

3c. Statistic for stability of sample ordering (year 1).

Given a panel of gene expression profiles derived from breast tumor samples, we typically have some information regarding patient clinical attributes including tumor grade, stage, ESR status, Her2 status, lymph node status, and ultimately patient outcome with respect to disease recurrence and overall survival. The canonical example of a question that is asked of such datasets is to identify molecular processes and/or cell types in the tumor that differ between patients of good and poor outcome. It is important to note that the assumption here is that tumors be broadly divided into these two groups before the analysis can be performed. Various bioinformatics tools like GSEA[8,14] exist for this type of analysis. However, the heterogeneity of breast cancer suggests that a simple a priori partition of the patients into classes such as good and bad outcome may not suffice. This is highlighted by the enormous differences that exist between subtypes, and the supposition that tumors of different subtypes recur for separate reasons. Indeed, previous attempts at identifying prognostic predictors of breast cancer outcome have largely been confounded by the subtypes, only having utility in a subset of patients[15]. Our observations suggest that the heterogeneity of breast cancers does not allow such a simple dichotomy, and it is nearly impossible to define 2 or more such classes a priori. Moreover, existing tools such as GSEA have a limitation in that they assume that a process is significantly differentially modulated between the bipartition of the patients. That is, these tools look for sets of genes with high expression in one category but low expression in the other. We argue that it is more natural for samples to display a range of activation levels for a given signature. This is a biological reality that is accepted within the community, but often ignored by bioinformatics methodologies. For example, it is common for Her2 to be genomically amplified one or more times in breast tumor cells, and its gene expression and membrane protein levels increase continuously in accordance. This increase has been directly linked to a corresponding change in signaling downstream of the receptor[16]. Staining of Her2 by immunohistochemistry (IHC) reveals a continuous range of intensities, which are scored from 0-3+ for simplicity, and often further reduced to simply Her2- or Her2+. While tumors are often summarized by a simple discretization, it is more natural for human breast tumors to display a range in signal activation levels or in the amount of various cell types present; bioinformatics methodologies should reflect this reality.

To overcome this problem, we have designed an intuitive approach that linearly orders tumors over individual signatures (Figure 1), thus measuring the strength of the particular response or cell type within the transcriptional profile of a tumor. Furthermore, in contrast to other traditional methodologies, our approach does not require *a priori* that tumors be binned into distinct classes. As such, the tool allows us to investigate continuous trends across the data, assessing the relative activation of signatures across a panel of patients. Using statistical approaches we have additionally developed, such orderings can be measured for robustness and other assessments of quality.

Since thousands of signatures are being employed, and each one generates a unique patient ordering, we have further developed statistical tests to identify those signatures from this large set that display 'interesting' behavior. The definition of 'interesting' is largely dependent on the particular question being asked of the patient dataset. For example, given a transcriptional signature of ESR activation (that is, the gene set corresponding to transcripts that are differentially expressed when ESR is over-expressed), patients are ordered according to their increasing relative expression of the signature. We may then ask whether the patient order is consistent with other assays for assessing the degree of ESR activity, including for instance IHC staining of the ESR protein (Figure 1). Alternatively, a signature may order patients in such a way that associations can be made with a variety of other histopathological/clinical parameters, such as tumor subtype or patient outcome. The development of statistics to identify such associations is not trivial. For example, in determining an association with patient outcome, the tumor ranks could be treated as a continuous variable under Cox regression, essentially asking whether an increase in patient rank linearly corresponds to a change in patient outcome. Alternatively, the patients on either end of the ordering may share good prognosis, with the patients in the center of the ordering having poor outcome. Both scenarios present relevant information about how a process or cell type relates to patient prognosis, but they require different means of analysis. There are benefits and drawbacks to the various approaches, and ultimately, any biological conclusions depend on such choices.

We have successfully developed a variety of statistics that are able to determine associations between the patient ordering and discrete clinical variables (such as ESR status or tumor subtype), continuous variables (such as age), as well as patient outcome. In addition, we have developed statistics that measure the stability of a patient ordering generated by a particular signature, when compared against the stability generated by a random set of genes. This allows us to filter out those signatures that are less trustworthy in the data.

The type of statistic described thus far treats each signature independently. However, a natural question arises as to whether dependencies exist between the patient orderings generated by each signature. There may be technical reasons for dependencies between signatures (e.g. they have many genes in common), or there may be some underlying biological reason. For such a set of signatures that order patients in a similar way, we wish to investigate whether they also tend to share associations with histological/clinical parameters and/or functional ontologies. To investigate this, we begin by calculating the correlation between every pair of patient orderings, and use this information to build a graph network with edges placed between nodes (signatures) that have a high correlation (figure 2). Highly interconnected regions of the graph are investigated for overrepresentations in associations with available histological/clinical parameters. This is not simply a technical investigation, but one with biological and clinical worth. The fact that processes are correlated tells us about how tumor cells respond to stress, and hints at the molecular level regulatory interactions that take place in tumor progression.

This in turn suggests better stratification of patients for the development and success of new treatment targets. Thus, such a signature-network approach identifies functionally-related signatures, even when the signatures represent different biological processes that share little or no genes in common.

Task 4. Application of framework to datasets (year 1-3):
4a. Apply signatures to human tumor datasets (year 1-2).
Milestone #2 Publication (year 2).

Our linear ordering procedure has been repeated for signatures within our catalogue BreSAT-DB, across a compendium of ~400 ductal carcinoma in situ and ~2000 invasive breast carcinomas, for which clinically annotated whole tumor gene expression data was available. Appropriate tests were used to identify statistically significant associations between the patient ordering generated by each signature, and histopathological/clinical variables including intrinsic subtype, ESR status, Her2 status, lymph node status, grade, recurrence, and overall survival. An interesting early finding was that the large majority of signatures have a significant association with certain clinical variables, such as ESR status and the tumor subtype. In fact, even random sets of genes tended to produce significant associations. This is a testament to the enormous transcriptional perturbations that occur downstream of specific molecular events, including activation of ESR. To compensate for this trend, the significance of an association with a given molecular signature is adjusted by resampling 10,000 random gene sets of the same size.

After adjustment, there remained a large number of signatures consistently having significant associations with the variables tested, and there was a surprising overlap in the signatures that associate with any given variable. (figure 2,3). Thus far, we have identified 239 signatures that consistently had a significant association with molecular subtype in at least half of the datasets investigated (adjusted pvalue <= 0.05). Typically this association was the result of Luminal A and Basal tumors having vastly different patient ranks. In addition, 207 signatures were found to consistently have significant associations with ER status, 23 with lymph node status, 125 with disease recurrence, and 116 with overall survival (161 combined total for patient outcome). As expected, signatures designed to predict patient outcome in breast cancer patients were all highly significant in the majority of datasets. Remarkably, however, we have been able to identify signatures with consistent, significant associations to patient outcome, but having no such associations to any of the other variables tested. These are signatures that encompass a variety of processes, such a response to hypoxia, VEGF signaling, or activation of the complement immune system. Because such signatures operate independently of known histopathological/clinical parameters, they represent a unique class with prognostic value across all subtypes, which contrasts the types of predictors that are in clinical use[15]. This is an important milestone, because it identifies molecular markers that are determinants of outcome in breast cancer, but have remained unrecognized to date. The identification of such elements is essential for the

development of new classes of treatments. Furthermore, our methodology represents a fundamentally different way of characterizing breast tumors. Whereas traditional approaches segment patients into classes according to the expression of a small number of genes, BreSAT comprehensively identifies the entire set of pathways, processes, responses, and cell types that define the disease. This exhaustive cataloguing of the molecular differences between subtypes is providing a more refined understanding, clinically and molecularly, of the underlying biology of the disease.

As there is some indication that breast tumors of each intrinsic subtype represent distinct biological entities, our analysis was further extended to observe how signatures associate with histopathological/clinical variables within each individual subtype. BreSAT was applied in isolation to patient sets belonging to each of the five intrinsic subtypes, and statistical associations were determined as before. Interestingly, these results revealed that each subtype tends to favor its own set of signatures (and by extension, processes) that associate with patient outcome. The luminal A subtype contained the largest number of signatures that were associated with patient outcome (recurrence and/or overall survival), most of which ordered patients in a manner that was independent of ER status, LN status, and grade. In contrast, tumors belonging to the luminal B subtype had only 7 signatures consistently associated with patient outcome in at last half of the datasets tested. Surprisingly, 5 of these 7 were signatures derived to specifically predict outcome in breast cancer patient. This suggests that patients with luminal B tumors are especially good candidates for therapeutic decision-making through genomic predictors. Tumors within the ERBB2 and Basal subtypes also had a small number of associations between signatures and patient outcome (8 and 2 respectively), possibly due to the smaller sample size of these subtypes. These associations related to processes such as TGF-Beta and p21 in the ERBB2 subtype, and CK1 and mRNA processing in the Basal subtype. The disparities in the results are perhaps not surprising, as the patients with tumors belonging to different subtypes tend to receive different treatments for their disease. However, our results are particularly applicable as indicators of how and why current treatments fail in different subsets of breast cancer patients.

Such results support our hypotheses that breast tumors can be described by the activation/repression of various molecular signatures, which can act in parallel or orthogonally to a tumor's intrinsic subtype, and are a consequence of the complex mix of cell types within the tumor. To better understand the contribution of different cell types to breast tumor biology and disease outcome, we next applied BreSAT to a dataset containing microdissected epithelium and stroma tissue from matched breast tumors (figure 4). As before, statistical tests were used to identify associations between signatures and histopathological/clinical variables of interest. Because the process was performed in matching tumor epithelium and stroma, we were able to distinguish between signatures that are macroenvironmental (present in all compartments of the tumor) vs those that are microenvironmental (present either in epithelium or stroma, but not both). Furthermore, our results have

revealed that some subsets of patients display remarkably similar signature activation/repression in matched tumor epithelium and stroma, whereas other patient subsets are enriched in microenvironment-specific responses.

We are additionally investigating the types of dependencies that exist between signatures. By quantifying the correlation between all possible pairs of signature-derived patient orders, we identify functional associations between signatures, even when the signatures represent vastly different biological processes that share little or no genes in common. Our analysis indicates that although there is an overrepresentation of highly correlated signatures with a significant number of genes in common, there additionally exist many correlated signature pairs with no overlap. We identify many such distinct types of processes and cell types that appear to be highly correlated to one-another, and are currently examining ways of subdividing our collection of signatures into a core set of groups. The fact that many processes are co-modulated suggests methods for building more robust and accurate prognostic signatures, that encompass a broader range of clinically-relevant characteristics with highly resilient signals.

In year 3, we had an unexpected and unique opportunity develop to apply our BreSAT framework to a novel dataset being generated by our collaborators in Oslo, Norway. This dataset currently comprises mRNA, lincRNA, miRNA, and SNP profiles for non-invasive ductal carcinoma in situ (DCIS) and invasive ductal carcinoma (IDC), currently totaling ~270 profiles, although additional NGS profiles are being developed. One of the goals of this work was to identify molecular differences between non-invasive and invasive breast cancer, which may indicate potential mechanisms that drive disease progression.

We determined those genes that significantly differentiated our set of all DCIS tumors from all IDC tumors. Similarly, we used the BreSAT framework to identify those signatures that significantly differentiated samples in the same manner. However, in both of these types of analyses, we observed an odd trend – those genes and signatures that differentiated DCIS from IDC were highly associated with the intrinsic subtype. For example, tumors classified as having a normal-like subtype, regardless of whether they were invasive or not, were always ranked amongst DCIS samples. Additionally, these same genes and signatures tended to work better at differentiating ESR-positive DCIS from IDC (which make up the majority of the dataset), than they did at differentiating ESR-negative DCIS from IDC. Furthermore, BreSAT-DB contains ~20 signatures that had been previously categorized as associated with progression in breast cancer. These were applied to our data, and in nearly all cases the same trends were observed. To further verify our findings, we applied these previously described genes and signatures to other breast cancer datasets in our compendium that contained both non-invasive and invasive samples. Although none of these other available datasets were as large as ours, making it difficult to determine significance within those subtypes containing a smaller number of samples, we again observed similar trends.

To overcome this issue, we sought to identify those genes and signatures that differentiate DCIS from IDC individually within each subtype. Remarkably, the genes and signatures that we identify represent diverse processes for each subtype, with very little overlap between subtypes (Figure 5). The biologies identified here generally reflect changes in cellular adhesion and proliferation in the luminal A subtype, changes in the extracellular matrix and fibroblasts amongst the luminal B subtype, changes in cellular differentiation amongst the ERBB2 subtype, and various immunological changes amongst the basal subtype. For example, while basal DCIS samples displayed no activation of a Th1 adaptive immune cell response, basal IDC samples had a statistically higher level of this immune response (Figure 6). This trend was not observable among other subtypes, and thus may represent a basal-specific mechanism involved in disease progression from a non-invasive to an invasive state. Work on this project is continuing, with a future focus on integrating information between the various array platforms, and with validation currently underway using tissue microarray slides.

Task 5. Hypothesis-driven generation of model systems (year 2-3):
5a. Selection of appropriate cell lines and mouse models (year 3).
5b. Molecular engineering of models (year 3).
5c. Analysis of modification success (year 3).
Milestone #3 Publication (year 3).

Several hundred samples of various mouse models and cell lines of the disease have been collected and formatted into our compendium. Our linear ordering procedure has been repeated for all ~6500 signatures within our catalogue BreSAT-DB, thus identifying which models have repression or activation of processes of interest. Not surprisingly, the cell lines are largely reflective of primary breast tumors in terms of the patterns of signature activation. For example, ESR positive cell lines tend to display activation of various endocrine-related signatures, while ESR negative cell lines tend to display activation of signatures related to MAPK-induced proliferation. Nonetheless, cell lines differ from human tumors in the activation of various signatures. For example, ESR positive human tumors display activation of various signatures related to cellular adhesion and interaction with the cellular microenvironment, while ESR positive cell lines do not. This may be explained by differences in the physical environment of the two sample types. As changes in the breast microenvironment has been shown to have an effect on disease outcome, this points to a major component that is lacking with 2-dimensional serum-based cell line models.

Initial comparisons between human breast tumors and mouse models of the disease indicate similar trends; while individual models tend to share molecular components with particular human subtypes, the similarities are imperfect. For example, over all ~6500 gene sets in BreSAT-DB, the MMTV-Neu model has an activation pattern that is highly correlated with human luminal A tumors (Figure 7). Both MMTV-Neu murine tumors and human luminal A tumors present relatively

high levels of signatures representing E2F3 silencing and cell cycle arrest. However, luminal A tumors consistently demonstrate high activation of signatures relating to ESR and other endocrine pathways; a property that is not shared by MMTV-Neu mouse tumors. This is not surprising, given that human luminal A tumors tend to be ESR positive, while MMTV-Neu tumors are not. Furthermore, MMTV-Neu murine tumors display activation of various immune components that are not shared by human luminal A tumors (Figure 8). Together, this implies where the MMTV-Neu murine model could be used to test hypotheses and treatments within the human luminal A subtype, and equally of value, when it should not be used.

Previously, we had demonstrated that expression of activated MET in murine mammary epithelium induces the formation of tumors, with a basal-like phenotype in approximately 50% of cases[17]. These tumors arise after an extended period of latency, with a low penetrance, and do not contain mutations in p53. This is in contrast to human basal tumors, which are known to display frequent mutations in p53, along with changes in the downstream responses of p53, and is associated with a more aggressive disease. Although there is now a well-established role for MET in basal and triple-negative breast cancer[17-19], we further sought to improve our mouse model by pairing the expression of activated MET with conditional loss of p53. Tumors in these mice arose with a low period of latency, a high penetrance, and a more homogeneous, spindloid pathology. Gene expression, miRNA, and aCGH profiles were generated for these tumors, giving us the opportunity to apply our BreSAT framework to the model and determine how well it reflected human breast cancer.

Our results suggested that overall, these spindloid tumors faithfully reflected the human claudin-low subtype of breast cancer. Using our signatures database (BreSAT-DB), we utilized human and cross-species intrinsic signatures to identify that the spindloid tumors had expression profiles most similar to human claudin-low tumors. Similarly, mRNA and miRNA signatures that had been derived from human and specifically identify human claudin-low tumors were applied to our mouse data. Using our linear ordering methodology and our associated statistics, we identified that these signatures were highly associated with our murine spindloid tumors. Additionally, the genes that are in common or differ between our mouse models, human claudin-low tumors, and human claudin-low cell lines were compared against BreSAT-DB. This analysis highlighted pathways related to epithelial–mesenchymal transition, MET signaling, and immune infiltration as shared between the human disease and mouse model, but none of statistical significance as differing between them. Moreover, we were able to demonstrate that these tumors were highly addicted to MET, requiring it to maintain proliferation and survival. Together, our work has highlighted MET as a cancer driver in this model, and may help to identify breast cancer patients that would benefit from anti-MET therapies. This work has been published in a high-impact journal[9], and is additionally available in the appendices of this report.

Key Research Accomplishments

- Construction of a comprehensive and highly annotated signature database, specific to breast cancer (BreSAT-DB). This database currently holds ~6500 gene sets, a large proportion of which were developed in breast-related tissue.

- Collection and formatting of ~20,000 data samples relating to breast cancer (BreSAT-Compendium). These comprise primarily gene expression profiles of invasive ductal carcinoma, but additionally include other types of molecular high-throughput data, samples representing different stages of the disease, and samples representing models for the disease.

- The generation of various visual and statistical methodologies to apply signatures to the collected datasets, and to determine the significance of associations between pathways, processes, responses, or cell types, and available histopathological/clinical parameters.

- Application of our signatures to human datasets, testing for statistical associations and dependencies between signatures.

- Application of our signatures to murine and cell line models of breast cancer, using the developed statistical tests to determine which signatures are highly and consistently activated in individual models.

- Use of our framework (in combination with experimental validation) to determine that MET and loss of p53 synergize to form tumors that faithfully model the claudin-low subtype of breast-cancer. This work has been published in a high-impact peer-reviewed journal[9] (see appendix).

Reportable Outcomes

Publications

Knight JF*, Lesurf R*, Zhao H, Pinnaduwage D, Davis RR, Saleh SM, Zuo D, Naujokas MA, Chughtai N, Herschkowitz JI, Prat A, Mulligan AM, Muller WJ, Cardiff RD, Gregg JP, Andrulis IL, Hallett MT, Park M. Met synergizes with p53 loss to induce mammary tumors that possess features of claudin-low breast cancer. PNAS. 2013 Apr 2;110(14):E1301-10.

*Authors contributed equally to the work.

Presentations

Title: Molecular features of subtype-specific progression from ductal carcinoma in situ to early invasive breast cancer
Conference: 12th Annual McGill Workshop on Bioinformatics in Barbados: Modern Biomarkers in Breast Cancer
Location: Holetown, Barbados
Date: January 2013

Title: Integrated molecular profiles identify mechanisms of subtype-specific progression from ductal carcinoma in situ to early invasive breast cancer
Conference: Personalized Cancer Care (talk delivered by Therese Sørlie)
Location: Oslo, Norway
Date: September 2012

Title: Breast Signature Analysis Tool (BreSAT): a framework for investigating the molecular networks of breast cancer
Conference: Era of Hope
Location: Orlando, Florida
Date: August 2011

Title: Breast Signature Analysis Tool (BreSAT): a framework for investigating the molecular networks of breast cancer
Conference: 10th Annual McGill Workshop on Bioinformatics in Barbados: Systems Approaches in Translational Breast Cancer Research
Location: Holetown, Barbados
Date: January 2011

Posters

Title: Integrated molecular profiles identify mechanisms of subtype-specific progression from ductal carcinoma in situ to early invasive breast cancer.
Conference: Personalized Cancer Care
Location: Oslo, Norway
Date: September 2012

Title: Breast Signature Analysis Tool (BreSAT): a framework for investigating the molecular networks of breast cancer
Conference: Era of Hope
Location: Orlando, Florida
Date: August 2011

Title: Breast Signature Analysis Tool (BreSAT): a framework for investigating the molecular networks of breast cancer
Conference: RECOMB Computational Cancer Biology 2010
Location: Oslo, Norway
Date: June 2010

Collection and normalization of breast-related data (BreSAT-Compendium)
In total, our compendium now includes ~20,000 human patient samples with associated histopathological/clinical data. Our compendium has been stratified by stages of disease progression (e.g. normal tissue, DCIS, IDC, metastases, etc.), type of sample (e.g. whole tumor versus cell-specific tissue derived by laser capture microdissection), adjuvant and neoadjuvant treatments, and type of data (e.g. gene expression microarrays, aCGH, miRNA, NGS, etc.). Our group is additionally in the process of generating additional next generation sequencing profiles for use. The collection involves a rigorous process of normalization and harmonization. Clinical parameters must be carefully matched to determine, for example, whether recurrence is measured as a local or distant event that takes place in a common 5- or 10-year time frame. This ensures that clinical information is directly comparable from one dataset to the next, and allows us to develop automated tools for analyzing the data. While our focus has been on human data, we also have hundreds of high-throughput samples representing models for the disease, including murine tumors and human cell lines.

Annotated signature database (BreSAT-DB)
Collection, refinement, and annotation of ~6,500 available molecular signatures with features such as the species and tissue they were generated in, as well as their general category (e.g. whether they are used to define a particular cell type, biological response, or a broad prognostic response). Within each of these categories, the signatures are further sub-classified as appropriate (e.g. signatures that define biological responses are sub-classified into one of ten hallmarks of cancer[6]). While we have collected numerous available gene sets from public databases, we have additionally focused on obtaining signatures from the literature that were specifically generated in breast-related tissues. This ensures that our signature database, BreSAT-DB, comprehensively and accurately reflects those pathways, processes, responses, and cell types that are relevant to breast cancer.

Programming package in R for data analysis (BreSAT)

We have developed numerous computational methodologies to load breast-related high-throughput data, to filter and visualize signatures of interest in the data, and statistics to quantify the relevance of such applications. These functions have been coded in the R programming language with a flexible design that allows them to be used by other researchers with various data types. The code has been formatted as an R package to be released for free through bioconductor.

Website

Much of the BreSAT framework has been designed for use in R. However, the vast majority of breast cancer researchers don't have the technical skills necessary to use it in this format. Therefore we're in the process of designing a website that can access an R session and generate associated figures and statistics based on simple point-and-click commands. The website is currently being run on a powerful server that should be able to handle incoming traffic from multiple sources simultaneously.

Conclusion

The framework we have described is a novel and important step towards better understanding the underlying pathways, processes, responses, and cell types that influence breast cancer progression and outcome. Our data compendiums represent the largest effort we are aware of to collect high-throughput breast-related data in an appropriately formatted and clinically annotated fashion. Similarly, our signature collection BreSAT-DB, contains the largest signature collection known to us, is thoroughly annotated, and crucially, is highly specific to breast cancer. Work is nearing completion, and the framework is set for release as both an R package and an interactive website.

Our analysis with the BreSAT framework has allowed us to piece together the interplay between individual molecular signatures, and to better understand how this interplay affects the phenotype of breast cancer. Our methodology introduces a unique and intuitive semi-supervised approach to pathway analysis, and is robust when multiple disparate high-throughput datasets are used. Crucially, it represents an entirely different way of classifying the disease. Instead of relying on the 'loudest' molecular signals that make up the majority of a transcriptional profile, the status of subtle but important biological pathways are taken into account. BreSAT provides the community with the means to comprehensively determine the classes of responses that characterize individual tumors.

Our analysis of primary human tumors has identified numerous processes that influence disease progression and outcome. In a similar manner, we have applied our methodology to cell line and mouse models of the disease. This has allowed us to determine which models best reflect individual aspects and/or subgroups of the human disease, and in what ways the models are different than primary tumors. In one specific example, we have used our framework to identify that synergy between the MET oncogene and loss of p53 lead to a tumor phenotype that reflects the human claudin-low subclass of breast cancer. In combination with experimental validation, our work has highlighted MET as a cancer driver in this model, and may help to identify patients that would benefit from anti-MET therapies.

References

1. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lønning PE, Børresen-Dale AL, Brown PO, Botstein D. Molecular portraits of human breast tumours. Nature. 2000 Aug 17;406(6797):747-52.

2. Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Eystein Lønning P, Børresen-Dale AL. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A. 2001 Sep 11;98(19):10869-74.

3. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, Lønning PE, Brown PO, Børresen-Dale AL, Botstein D. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A. 2003 Jul 8;100(14):8418-23.

4. Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, He X, Perou CM. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res. 2010;12(5):R68.

5. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, Pietenpol JA. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest. 2011 Jul;121(7):2750-67.

6. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Gräf S, Ha G, Haffari G, Bashashati A, Russell R, McKinney S; METABRIC Group, Langerød A, Green A, Provenzano E, Wishart G, Pinder S, Watson P, Markowetz F, Murphy L, Ellis I, Purushotham A, Børresen-Dale AL, Brenton JD, Tavaré S, Caldas C, Aparicio S. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature. 2012 Apr 18;486(7403):346-52.

7. Culhane AC, Schwarzl T, Sultana R, Picard KC, Picard SC, Lu TH, Franklin KR, French SJ, Papenhausen G, Correll M, Quackenbush J. GeneSigDB--a curated database of gene expression signatures. Nucleic Acids Res. 2010 Jan;38(Database issue):D716-25.

8. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005 Oct 25;102(43):15545-50.

9. Knight JF, Lesurf R, Zhao H, Pinnaduwage D, Davis RR, Saleh SM, Zuo D, Naujokas MA, Chughtai N, Herschkowitz JI, Prat A, Mulligan AM, Muller WJ, Cardiff RD, Gregg JP, Andrulis IL, Hallett MT, Park M. Met synergizes with p53 loss to induce mammary tumors that possess features of claudin-low breast cancer. Proc Natl Acad Sci U S A. 2013 Apr 2;110(14):E1301-10.

10. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011 Mar 4;144(5):646-74.

11. Chang JT, Gatza ML, Lucas JE, Barry WT, Vaughn P, Nevins JR. SIGNATURE: a workbench for gene expression signature analysis. BMC Bioinformatics. 2011 Nov 14;12:443.

12. Kristensen VN, Vaske CJ, Ursini-Siegel J, Van Loo P, Nordgard SH, Sachidanandam R, Sørlie T, Wärnberg F, Haakensen VD, Helland Å, Naume B, Perou CM, Haussler D, Troyanskaya OG, Børresen-Dale AL. Integrated molecular profiles of invasive breast tumors and ductal carcinoma in situ (DCIS) reveal differential vascular and interleukin signaling. Proc Natl Acad Sci U S A. 2012 Feb 21;109(8):2802-7.

13. Vaske CJ, Benz SC, Sanborn JZ, Earl D, Szeto C, Zhu J, Haussler D, Stuart JM. Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. Bioinformatics. 2010 Jun 15;26(12):i237-45.

14. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstråle M, Laurila E, Houstis N, Daly MJ, Patterson N, Mesirov JP, Golub TR, Tamayo P, Spiegelman B, Lander ES, Hirschhorn JN, Altshuler D, Groop LC. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet. 2003 Jul;34(3):267-73.

15. Fan C, Oh DS, Wessels L, Weigelt B, Nuyten DS, Nobel AB, van't Veer LJ, Perou CM. Concordance among gene-expression-based predictors for breast cancer. N Engl J Med. 2006 Aug 10;355(6):560-9.

16. Hodgson JG, Malek T, Bornstein S, Hariono S, Ginzinger DG, Muller WJ, Gray JW. Copy number aberrations in mouse breast tumors reveal loci and genes important in tumorigenic receptor tyrosine kinase signaling. Cancer Res. 2005 Nov 1;65(21):9695-704.

17. Ponzo MG, Lesurf R, Petkiewicz S, O'Malley FP, Pinnaduwage D, Andrulis IL, Bull SB, Chughtai N, Zuo D, Souleimanova M, Germain D, Omeroglu A, Cardiff RD, Hallett M, Park M. Met induces mammary tumors with diverse histologies and is associated with poor outcome and human basal breast cancer. Proc Natl Acad Sci U S A. 2009 Aug 4;106(31):12903-8.

18. Graveel CR, DeGroot JD, Su Y, Koeman J, Dykema K, Leung S, Snider J, Davies SR, Swiatek PJ, Cottingham S, Watson MA, Ellis MJ, Sigler RE, Furge KA, Vande Woude GF. Met induces diverse mammary carcinomas in mice and is associated with human basal breast cancer. Proc Natl Acad Sci U S A. 2009 Aug 4;106(31):12909-14.

19. Dutta B, Pusztai L, Qi Y, André F, Lazar V, Bianchini G, Ueno N, Agarwal R, Wang B, Shiang CY, Hortobagyi GN, Mills GB, Symmans WF, Balázsi G. A network-based, integrative study to identify core biological pathways that drive breast cancer clinical subtypes. Br J Cancer. 2012 Mar 13;106(6):1107-16.

20. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, Witteveen A, Glas A, Delahaye L, van der Velde T, Bartelink H, Rodenhuis S, Rutgers ET, Friend SH, Bernards R. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med. 2002 Dec 19;347(25):1999-2009.

21. Abba MC, Hu Y, Sun H, Drake JA, Gaddis S, Baggerly K, Sahin A, Aldaz CM. Gene expression signature of estrogen receptor alpha status in breast cancer. BMC Genomics. 2005 Mar 11;6:37.

22. Finak G, Sadekova S, Pepin F, Hallett M, Meterissian S, Halwani F, Khetani K, Souleimanova M, Zabolotny B, Omeroglu A, Park M. Gene expression signatures of morphologically normal breast tissue identify basal-like tumors. Breast Cancer Res. 2006;8(5):R58.

23. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000 Jan 1;28(1):27-30.

24. Chin K, DeVries S, Fridlyand J, Spellman PT, Roydasgupta R, Kuo WL, Lapuk A, Neve RM, Qian Z, Ryder T, Chen F, Feiler H, Tokuyasu T, Kingsley C, Dairkee S, Meng Z, Chew K, Pinkel D, Jain A, Ljung BM, Esserman L, Albertson DG, Waldman FM, Gray JW. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer Cell. 2006 Dec;10(6):529-41.

25. Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, Rasmussen KE, Jones LP, Assefnia S, Chandrasekharan S, Backlund MG, Yin Y, Khramtsov AI, Bastein R, Quackenbush J, Glazer RI, Brown PH, Green JE, Kopelovich L, Furth PA, Palazzo JP, Olopade OI, Bernard PS, Churchill GA, Van Dyke T, Perou CM. Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. Genome Biol. 2007;8(5):R76.

26. Olsson AY, Feber A, Edwards S, Te Poele R, Giddings I, Merson S, Cooper CS. Role of E2F3 expression in modulating cellular proliferation rate in human bladder and prostate cancer cells. Oncogene. 2007 Feb 15;26(7):1028-37.

27. Harvell DM, Richer JK, Singh M, Spoelstra N, Finlayson C, Borges VF, Elias AD, Horwitz KB. Estrogen regulated gene expression in response to neoadjuvant endocrine therapy of breast cancers: tamoxifen agonist effects dominate in the presence of an aromatase inhibitor. Breast Cancer Res Treat. 2008 Dec;112(3):489-501.

28. Rody A, Holtrich U, Pusztai L, Liedtke C, Gaetje R, Ruckhaeberle E, Solbach C, Hanker L, Ahr A, Metzler D, Engels K, Karn T, Kaufmann M. T-cell metagene predicts a favorable prognosis in estrogen receptor-negative and HER2-positive breast cancers. Breast Cancer Res. 2009;11(2):R15.

Met synergizes with p53 loss to induce mammary tumors that possess features of claudin-low breast cancer

Jennifer F. Knight[a,1], Robert Lesurf[a,b,1], Hong Zhao[a], Dushanthi Pinnaduwage[c], Ryan R. Davis[d], Sadiq M. I. Saleh[a,b], Dongmei Zuo[a], Monica A. Naujokas[a], Naila Chughtai[a], Jason I. Herschkowitz[e], Aleix Prat[f,g], Anna Marie Mulligan[h,i], William J. Muller[a,b], Robert D. Cardiff[d], Jeff P. Gregg[d], Irene L. Andrulis[c,h,i,j], Michael T. Hallett[a], and Morag Park[a,b,k,2]

[a]Goodman Cancer Research Centre, McGill University, Montreal, QC, Canada H3A 1A3; [b]Department of Biochemistry, McGill University, Montreal, QC, Canada H2W 1S6; [c]Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, ON, Canada M5G 1X5; [d]Center for Comparative Medicine, University of California, Davis, CA 95616; [e]Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, TX 77030; [f]Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC 27514; [g]Vall d'Hebron Institute of Oncology, 08035 Barcelona, Spain; [h]Department of Laboratory Medicine, St. Michael's Hospital, Toronto, ON, Canada M5B 1W8; [i]Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada M5S 1A1; [j]Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada M5S 1A8; and [k]Department of Oncology, McGill University, Montreal, QC, Canada H2W 1S6

Edited by Tak W. Mak, The Campbell Family Institute for Breast Cancer Research, Ontario Cancer Institute at Princess Margaret Hospital, University Health Network, Toronto, ON, Canada, and approved February 13, 2013 (received for review June 18, 2012)

Triple-negative breast cancer (TNBC) accounts for ~20% of cases and contributes to basal and claudin-low molecular subclasses of the disease. TNBCs have poor prognosis, display frequent mutations in tumor suppressor gene p53 (*TP53*), and lack targeted therapies. The MET receptor tyrosine kinase is elevated in TNBC and transgenic *Met* models (Met^{mt}) develop basal-like tumors. To investigate collaborating events in the genesis of TNBC, we generated Met^{mt} mice with conditional loss of murine p53 (*Trp53*) in mammary epithelia. Somatic *Trp53* loss, in combination with Met^{mt}, significantly increased tumor penetrance over Met^{mt} or *Trp53* loss alone. Unlike Met^{mt} tumors, which are histologically diverse and enriched in a basal-like molecular signature, the majority of Met^{mt} tumors with *Trp53* loss displayed a spindloid pathology with a distinct molecular signature that resembles the human claudin-low subtype of TNBC, including diminished claudins, an epithelial-to-mesenchymal transition signature, and decreased expression of the microRNA-200 family. Moreover, although mammary specific loss of *Trp53* promotes tumors with diverse pathologies, those with spindloid pathology and claudin-low signature display genomic *Met* amplification. In both models, MET activity is required for maintenance of the claudin-low morphological phenotype, in which MET inhibitors restore cell-cell junctions, rescue claudin 1 expression, and abrogate growth and dissemination of cells in vivo. Among human breast cancers, elevated levels of MET and stabilized TP53, indicative of mutation, correlate with highly proliferative TNBCs of poor outcome. This work shows synergy between MET and TP53 loss for claudin-low breast cancer, identifies a restricted claudin-low gene signature, and provides a rationale for anti-MET therapies in TNBC.

Met RTK | EMT | mouse model | gene expression

Despite recent improvements in breast cancer mortality, this disease remains the second leading cause of cancer-related deaths for women worldwide (1). Gene expression profiling and molecular pathology have revealed that breast cancers naturally divide into luminal A and B, human epidermal growth factor receptor 2 (HER2)-enriched, basal-like, and the recently identified claudin-low subtypes (2, 3). Targeted therapies that rely on tumor cell expression of estrogen and v-erb-b2 erythroblastic leukemia viral oncogene homolog 2 (ErbB2) receptors can be effective in the treatment of luminal and HER2-positive breast cancers (4). However, basal-like and claudin-low breast cancers are predominately negative for these receptors, referred to as triple negative (TN), and are associated with poor prognosis. TN breast cancers account for up to 20% of breast cancer cases (5), emphasizing the need to identify molecular targets for their treatment.

Claudin-low tumors were originally distinguished from other subtypes on the basis of gene expression profiling (3) and have subsequently been correlated with tumors of metaplastic and medullary pathology (6). These tumors are characterized by loss of tight junction markers (notably claudins) and high expression of markers of epithelial-to-mesenchymal transition (EMT), in addition to being enriched for markers of mammary stem cells (6).

Signaling through MET, the receptor tyrosine kinase (RTK) for hepatocyte growth factor (HGF) influences diverse cellular processes during both developmental and cancer progression (7, 8). MET is expressed in the epithelium of numerous tissues, including breast, and regulates cell proliferation, migration, and invasion, as well as EMT (7, 8). Increased expression of MET is associated with TN breast cancers and correlates with poor outcome (8–11). In normal breast, activation of MET in ductal epithelium can occur through paracrine signaling, as a result of the secretion of HGF by stromal fibroblasts, and increased amounts of HGF are detected in serum of patients with breast cancer who have high-grade disease (12, 13).

Transgenic mice expressing a weakly oncogenic variant of *Met* under the control of the murine mammary tumor virus (MMTV)

Significance

Triple-negative breast cancers lack targeted therapies and are subdivided into molecular subtypes, including basal and claudin-low. Preclinical models representing these subtypes are limited. We have developed a murine model in which mammary gland expression of a receptor tyrosine kinase (MET) and loss of tumor suppressor gene p53 (*Trp53*), synergize to promote tumors with pathological and molecular features of claudin-low breast cancer. These tumors require MET signaling for proliferation, as well as mesenchymal characteristics, which are key features of claudin-low biology. This work associates MET expression and p53 loss with claudin-low breast cancers and highly proliferative breast cancers of poor outcome.

Author contributions: J.F.K., R.L., I.L.A., M.T.H., and M.P. designed research; J.F.K., R.L., H.Z., D.P., R.R.D., S.M.I.S., D.Z., M.A.N., and N.C. performed research; J.F.K., R.L., D.P., R.R.D., S.M.I.S., J.I.H., A.P., A.M.M., and R.D.C. analyzed data; W.J.M. and J.P.G. contributed new reagents/analytic tools; and J.F.K., R.L., I.L.A., M.T.H., and M.P. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE41748).

[1]J.F.K. and R.L. contributed equally to this work.

[2]To whom correspondence should be addressed. E-mail: morag.park@mcgill.ca.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1210353110/-/DCSupplemental.

promoter (MMTV-Met^{mt}), or knock-in of Met^{mt} into its endogenous promoter, develop mammary tumors that are histologically diverse (14, 15). Consistent with elevated MET in TN breast cancer, 50% of MMTV-Met^{mt} tumors exhibit a molecular signature of the basal-like subclass of human breast cancer and are positive for basal cytokeratins (14, 15). However, the long latency of the MMTV-Met^{mt} model supports the requirement for cooperating oncogenic events. Loss-of-function mutations in the tumor suppressor gene *TP53* (tumor protein p53) are detected in ~80% of TN breast cancers (2). Interplay between TP53 and MET is supported by the observation that in a mouse model of mammary tumorigenesis involving *Trp53* (murine p53) deletion, 73% of tumors carry amplification of *Met* (16). Moreover, *Met* mRNA levels are regulated by the p53-regulated microRNA (miRNA) miR34a (17). However, synergy between MET and *Trp53* loss during mammary tumor formation has not been tested.

To study the consequences of *Trp53* loss during MET-induced mammary tumorigenesis, we generated a conditional mouse model in which mammary gland–specific expression of *Met* (MMTV-Met^{mt}) is combined with Cre-recombinase (MMTV-Cre)–mediated deletion of floxed *Trp53* alleles in the mammary gland. We document a significant reduction in tumor latency coupled with a dramatic increase in tumor penetrance in MMTV-Met^{mt};Trp53fl/+;Cre mice compared with MMTV-Met^{mt} and a significant increase in penetrance compared with Trp53fl/+;Cre mice. The majority of mammary tumors that arise in MMTV-Met^{mt};Trp53fl/+;Cre mice and Trp53fl/+;Cre mice possess a distinctive spindloid pathology, and a comparison of gene expression data with human breast cancer datasets reveals a significant correlation between these mammary tumors and human claudin-low breast cancer. In both cases, the claudin-low phenotype is correlated with amplification of *Met* and requires continuous MET signaling. This work highlights the fact that MET and TP53 loss act synergistically in promoting breast tumors and provides a model to study the claudin-low subtype.

Results

MMTV-Met^{mt};Trp53fl/+;Cre Tumors Exhibit a Predominately Spindloid Pathology. To investigate the consequence of elevated MET in the absence of functional TP53, we generated a transgenic mouse model in which mammary gland expression of a weakly oncogenic MET receptor (MMTV-Met^{mt}) is combined with conditional deletion of *Trp53* in the mammary glands of FVB/N [Friend Leukaemia virus type B (susceptibility)-NIH] mice (MMTV-Met^{mt};Trp53fl/+;MMTV-Cre-recombinase). Compared with MMTV-Met^{mt} or Trp53fl/+;Cre control mice, we observed a dramatic increase in tumor penetrance, going from 31% and 24%, respectively, to 70% for MMTV-Met^{mt};Trp53fl/+;Cre mice (Table 1 and Fig. 1*A*). Moreover, although the MMTV-Met^{mt} model required multiple rounds of pregnancy to stimulate tumor development, 71% of virgin MMTV-Met^{mt};Trp53fl/+;Cre mice developed tumors (Table 1). Unlike the MMTV-Met^{mt} model, in which a spectrum of tumor pathologies was observed (14), the majority of mammary tumors that arose in MMTV-Met^{mt};Trp53fl/+;Cre mice (80%) and, to a lesser extent, in Trp53fl/+;Cre mice (63%) displayed a spindloid pathology, with the remaining tumors being poorly differentiated adenocarcinomas (Fig. 1*B*).

Cytokeratin (CK) expression can be used to infer the differentiation status of breast tumors (17, 18). Interestingly, although nonspindloid MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre adenocarcinomas expressed basal (CK14) and luminal (CK8/18) cytokeratins, as well as CK5 (associated with progenitor cells), spindloid tumors showed only weak and sporadic expression of all CKs tested (CK14, 8/18, 5/6) (Fig. S1*A*). Spindloid tumor cells stained strongly for the mesenchymal marker vimentin and were negative for the epithelial marker E-cadherin (Fig. S1), which is supportive of an EMT (20). Interestingly, coexpression of both cytokeratins and vimentin was detected by immunofluorescence in spindloid tumor cells as well as hyperplastic glands (Fig S1*B*), thus capturing EMTs. Together, these data support the idea that expression of activated MET in combination with the loss of *Trp53* in the mouse mammary gland promotes the formation of tumors with high penetrance and pronounced features that are typical of EMT.

MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre Tumors Undergo Loss of Heterozygosity for *Trp53* and Selectively Amplify the Endogenous *Met* Locus. Models of mammary tumorigenesis involving loss of a single allele of a tumor suppressor gene frequently undergo loss of heterozygosity during tumor progression, resulting in loss of the second allele (21). Consistent with this, all MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre mammary tumors tested showed Cre-mediated deletion of the conditional *Trp53* allele as well as loss of the wild-type (unfloxed) *Trp53* allele (Fig. S2). As loss of TP53 is associated with genomic instability (22), we used array-based comparative genomic hybridization (aCGH) to investigate whether consistent chromosomal alterations were associated with the MMTV-Met^{mt};Trp53fl/+;Cre and/or Trp53fl/+;Cre tumors. In addition to validating loss of the *Trp53* locus (Fig. S3*C*), array-CGH data also showed copy number changes consistent with human breast cancer. For example, three of seven MMTV-Met^{mt};Trp53fl/+;Cre spindloid tumors (but not Trp53fl/+;Cre spindloid tumors) showed gain of the locus encoding myelocytomatosis oncogene (*Myc*) (MsChr15:61.8Mb) (Fig. S3), which is amplified in 46.7% of human TN breast cancers of the claudin-low subclass (23). Although *Myc* amplification was not detected in Trp53fl/+;Cre spindloid tumors, both MMTV-Met^{mt};Trp53fl/+;Cre tumors and Trp53fl/+;Cre tumors with a spindloid component contained genomic amplification of the endogenous *Met* locus (Chr6 17.4–17.5Mb) (Fig. 1*C* and Fig. S3). Although variable, tumors contained a broad region of amplification at this locus (Chr6 16.7–18.2Mb), which included not only *Met* but also other genes adjacent to *Met*; including *Cav1* (caveolin 1), *Cav2* (caveolin 2), *Wnt2* (wingless-related MMTV-integration site 2) and *Cftr* (cystic fibrosis transmembrane conductance regulator) (Fig. S3). Notably, amplification of *Met* was absent in all Trp53fl/+;Cre tumors of adenocarcinoma pathology. The association between Met amplification and *Trp53*-null mammary tumors of spindloid but not adenocarcinoma-type pathology is highly significant ($P = 0.01786$), supporting an association between *Met* amplification and *Trp53*-deficient tumors with spindle-cell pathology.

Consistent with *Met* amplification, MMTV-Met^{mt};Trp53fl/+;Cre tumors showed strong immunohistochemical staining for the endogenous murine MET protein (Fig. 1*D*). In tumors as well as tumor lysates, the murine MET protein was highly phosphorylated on tyrosines 1234/5 (within the activation loop), consistent with its amplification and constitutive activation (Fig. 1*D* and Fig. S4) (6). This supports a possible "addiction" of the tumors to MET signaling. Endogenous *Met* amplification in MMTV-Met^{mt};Trp53fl/+;Cre tumors correlated with repression of the MMTV-Met^{mt} transgene (Fig. 1*D* and Fig. S4) and is consistent with suppression of the MMTV promoter after EMT, as shown previously (24). Notably, Trp53fl/+;Cre spindloid tumors, but not adenocarcinomas, also expressed elevated levels of endogenous murine MET at similar levels of activity to that of MMTV-Met^{mt};Trp53fl/+;Cre tumors (Fig. S4). Thus, genomic amplification of *Met* leads to constitutive activation of the MET RTK in the absence of its ligand HGF, supporting a potential dependency of these *Trp53*-deficient mammary tumors on MET signaling.

MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre Spindloid Tumors Are Characterized by a Strong EMT, Met Signaling Axis, and Significant Immune Infiltrate. To gain insight into the contribution of *Trp53* loss to *Met*-induced mammary tumorigenesis, gene expression profiles were generated from 14 MMTV-Met^{mt};Trp53fl/+;Cre, 8 Trp53fl/+;Cre tumors, 8 MMTV-Met^{mt} tumors, and 11 whole mammary gland (mammary fat pad, MFP) controls. Unsupervised hierarchical clustering with those genes that have an interquartile range greater than or equal to 2 over all samples identified three distinct clusters (Fig. 2*A*). The clusters were associated with tumor pathology in which all MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre spindloid tumors clustered together and tumors with an ade-

Table 1. Tumor penetrance and latency values for mammary tumor development in MMTV-Met^{mt};Trp53fl/+;Cre, MMTV-Met^{mt} and Trp53fl/+;Cre mice

Parity	Genotype	Tumor-bearing mice/ total mice	Penetrance, %	Latency, d
Nulliparous	MMTV-Met^{mt};Trp53fl/+;Cre	15/21	71.4	278
	Trp53fl/+;Cre	4/12	33.3	305
Multiparous	MMTV-Met^{mt};Trp53fl/+;Cre	13/19	68.4	280
	Trp53fl/+;Cre	2/13	15	276
	MMTV-Met^{mt}	16/52	31	430

Loss of mammary gland expression of *Trp53* in the MMTV-Met^{mt} model led to an increase in tumor penetrance and shortened latency, in addition to abrogating the requirement for parity for tumor development. Compared with Trp53fl/+;Cre control mice, MMTV-Met;Trp53fl/+;Cre mice developed tumors with a similar latency but at a significantly higher penetrance, indicating *Met* expression as an important event in tumor initiation.

nocarcinoma pathology clustered together, regardless of genotype. Normal mammary gland controls formed a distinct cluster away from the tumor samples. Genes differentially expressed between clusters are indicated in Dataset S1, Tables S1–S3.

Compared with MMTV-Met^{mt} tumors or normal MFP (Dataset S1, Tables S1–S3), a striking feature of MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre spindloid tumors was high expression of several markers of the previously determined EMT core signature (*Snai1/2*, *Twist1/2*, and *Zeb1/2*) (Fig. 2 *B* and *C*) (25), weak expression of cytokeratins as observed by immunohistochemical (IHC) analysis (Fig. S1 and Fig. 2*B*), and decreased representation of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) processes such as cell-cell junction organization, tight junction, and cell junction maintenance (Fig. 2*B* and Dataset S1, Tables S4–S7).

Analysis of the genes differentially expressed between MMTV-Met^{mt};Trp53fl/+;Cre spindloid and MMTV-Met^{mt} tumors also identified enrichment for GO and KEGG categories such as actin filament–based movement and regulation of cell projection organization (Dataset S1, Table S4 and Fig. 2*B*), as well as inflammatory response, positive regulation of macrophage chemotaxis, regulation of lymphocyte-mediated immunity, cytokine-cytokine receptor interaction, and chemokine signaling pathway (Dataset S1, Table S5). Consistent with this, high expression of several chemokines and chemokine receptors associated with monocyte and lymphocytic infiltration (*Ccr1*, *Cxcl10*, and *Cxcl1*) (Dataset S1, Table S2) (26, 27) suggested a strong inflammatory response in MMTV-Met^{mt};Trp53fl/+;Cre tumors. Immunostaining for the T- and B-lymphocyte markers CD3 and CD20 (Fig. S5 *A* and *B*) and the macrophage marker F4/80 (Fig. S5*C*) revealed elevated lymphocytic and macrophage content in MMTV-Met^{mt};Trp53fl/+;Cre spindloid tumors compared with in MMTV-Met^{mt} tumors.

In addition, the GO analysis included the category HGF receptor signaling pathway, reflecting a strong MET signaling axis

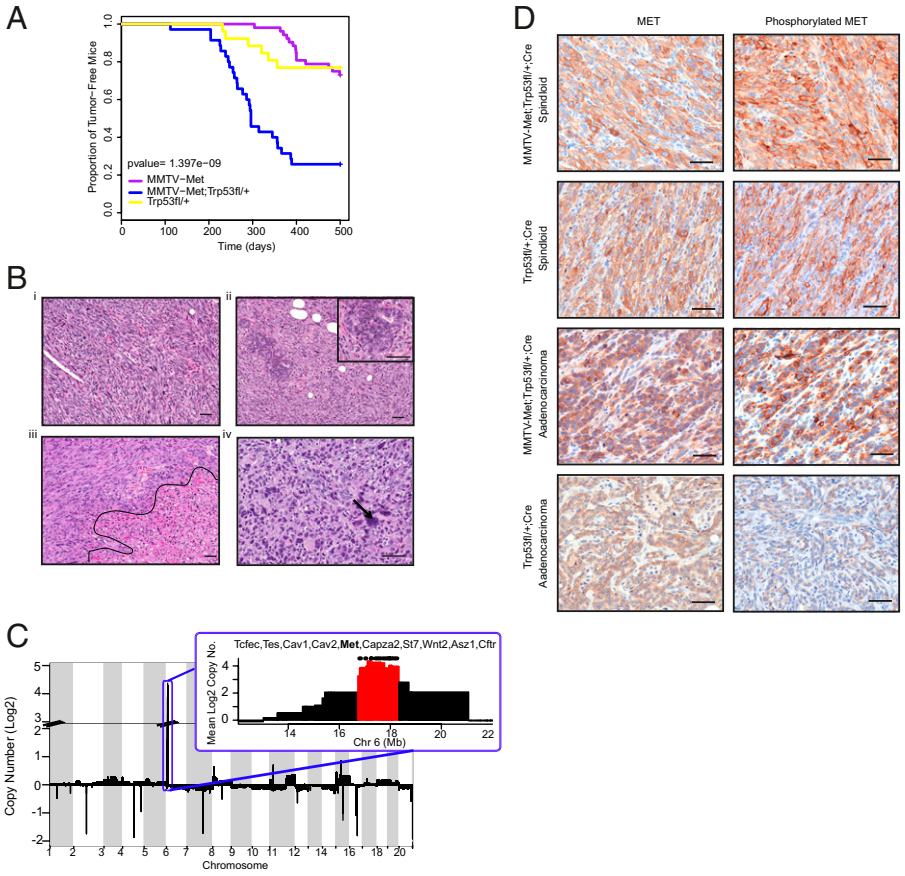


Fig. 1. MMTV-Met^{mt};Trp53fl/+;Cre mammary tumors are highly penetrant, have a spindloid pathology, and selectively amplify the endogenous *Met* locus. A Kaplan-Meier plot illustrates that MMTV-Met^{mt};Trp53fl/+;Cre (n = 35) and Trp53fl/+;Cre mice (n = 25) have similar tumor onsets (~300 d), occurring earlier than tumors in MMTV-Met^{mt} mice (n = 52) (~400 d) (*A*). However, MMTV-Met^{mt};Trp53;Cre mice are associated with a significantly higher tumor penetrance (~70%) compared with Trp53fl/+;Cre mice (~24%), resulting in a steeper curve (*A*). Tumor pathology was similar between MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre mice, ranging from spindloid to poorly differentiated adenocarcinomas (*B*). Cells with enlarged nuclei (arrow in *B*, *iv*) and large areas of necrosis (outlined in *B*, *iii*) were common. Spindloid tumors often contained ducts with atypical morphology (*Inset*, *B*, *ii*). All MMTV-Met^{mt};Trp53fl/+;Cre tumors contained genomic amplification of *Met* and adjacent loci, as determined by array-CGH (*C*), a phenomenon also observed in Trp53fl/+Cre tumors of spindloid pathology but not in Trp53fl/+;Cre adenocarcinomas (Fig. S2). High expression and activation (phosphorylation) of endogenous MET in MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre tumors was confirmed by immunostaining (*D*). A Trp53fl/+;Cre adenocarcinoma without amplification of *Met* and little activated MET is shown as a comparison (*D*). (Scale bars, 50 μm.)

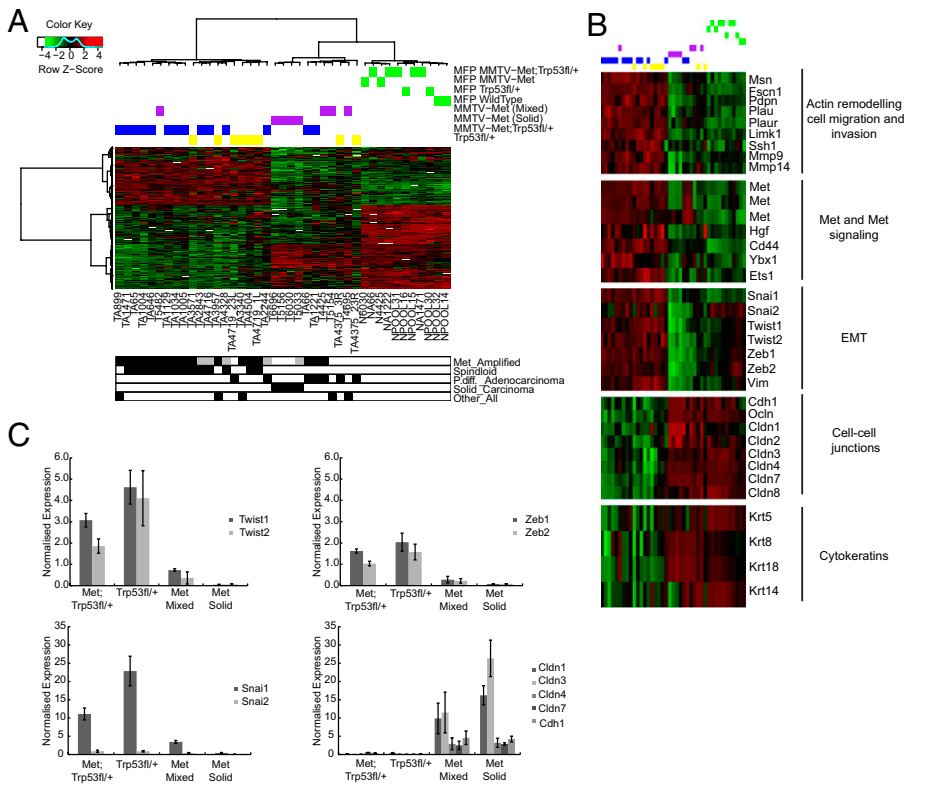


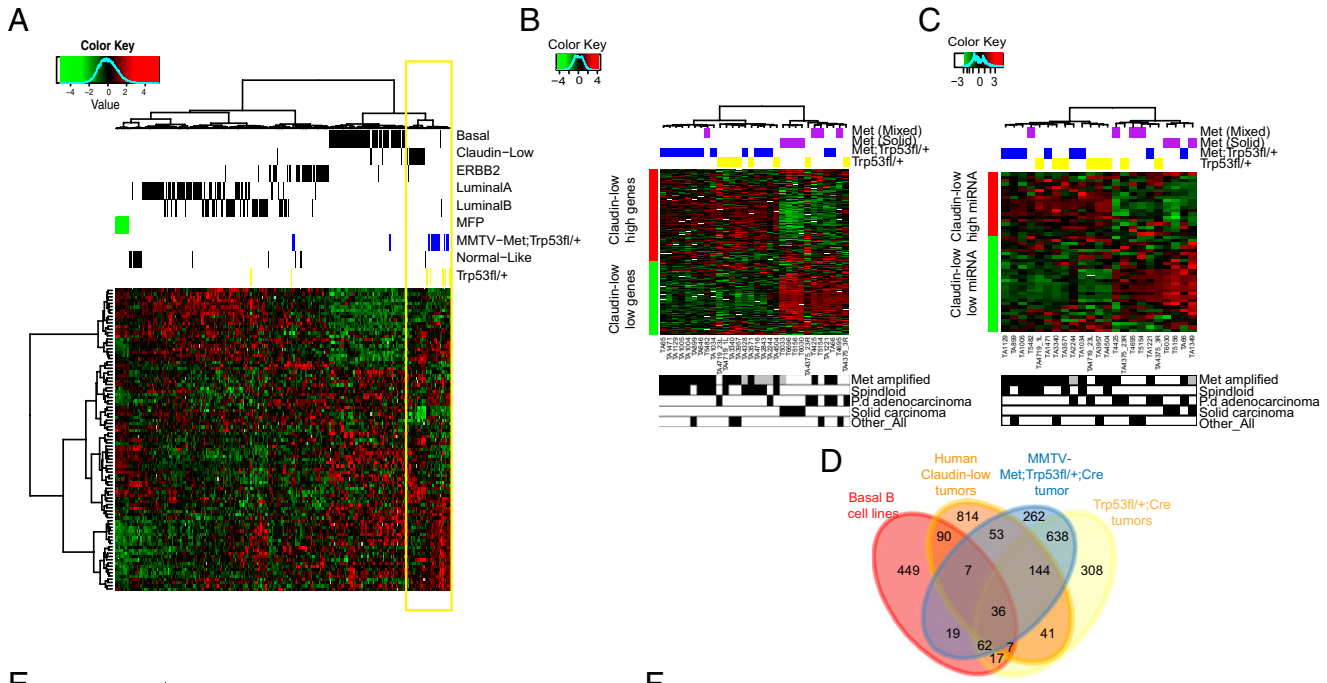
Fig. 2. MMTV-Met^{mt};Trp53fl/+;Cre spindloid tumors display elevated expression of genes associated with a mesenchymal, migratory phenotype and are distinct from MMTV-Met^{mt} mammary tumors. Unsupervised hierarchical clustering identifies three distinct groups. In the first group, 12 MMTV-Met^{mt};Trp53fl/+;Cre tumors (blue) form a cluster with six Trp53fl/+;Cre tumors (yellow) and one MMTV-Met^{mt} tumor (purple); this cluster represents tumors of predominantly spindloid pathology and with genomic amplification of *Met*. In the next cluster, poorly differentiated adenocarcinomas (two MMTV-Met^{mt};Trp53fl/+;Cre and two Trp53fl/+;Cre tumors) cluster with tumors of the MMTV-Met^{mt} model. MMTV-Met^{mt} tumors further segregate into solid and mixed subtypes in accordance with their pathology (14). Normal mammary gland controls (green) form the third cluster. Tumor characterizations below the heat map are represented in white for negative, black for positive, and gray for unknown. "Other_All" refers to tumors of various pathology types; for example, tumor A899 contained regions of spindloid and adenocarcinoma-type pathologies. Genes highly expressed in MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre spindloid tumors are associated with cell migration and invasion, signaling through the MET receptor, and EMT (*B*). Low expression of cell-cell junction markers and moderate expression of epithelial cytokeratins is also observed (*B*). A number of these genes were validated by qRT-PCR (n = 5 MMTV-Met^{mt};Trp53fl/+;Cre, 5 Trp53f/+;Cre tumors, 3 MMTV-Met^{mt} mixed tumors, and 3 MMTV-Met^{mt} solid tumors); expression relative to wild-type mammary gland is shown. Error bars, SEM (*C*).

within MMTV-Met^{mt};Trp53fl/+;Cre tumors (Dataset S1, Table S4). Consistent with *Met* amplification and activation, both MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre spindloid tumors show elevated expression of the *Met* gene, in addition to high expression of the MET receptor ligand *Hgf*, *Cd44* (a potential coreceptor for MET) (28), *Ets1*, and *Ybx1* (proposed transcriptional activators of *Met*) (Fig. 2*B* and Dataset S1, Tables S1–S3) (29, 30).

MMTV-Met^{mt};Trp53fl/+;Cre Tumors and Trp53fl/+;Cre Spindloid Tumors Cluster with the Claudin-Low Subtype of TN Breast Cancers. To determine whether MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre tumors were representative of a subtype of human breast cancer, gene expression profiles were compared with those of Herschkowitz and colleagues (3). Notably, all MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre tumors of spindloid, but not adenocarcinoma, pathology clustered with the claudin-low subclass of human breast cancers (Fig. 3*A*). The human claudin-low subclass signature reflects high expression of transcriptional drivers of EMT and low expression of markers of adherens and tight junctions, such as E-cadherin and claudins 1, 3, 4, and 7 (6). As validated by quantitative RT-PCR, MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre spindloid tumors showed similar expression of genes within this signature, expressing high levels of *Snai1/2*, *Twist1/2*, and *Zeb1/2* (Fig. 2*C*) and low levels of claudins such as *Cldn1,3,4* and *7* and *E-cadherin* (Fig. 2*C*). Importantly, application of a claudin-low subclass gene signature derived from human tumors (6) identified MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+Cre spindloid tumors as strongly correlative ($P <$ 0.0001) (Fig. 3*B*). Conversely, application of the differentially expressed gene signature from MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre spindloid tumors to human breast cancer subtypes induced a cluster of claudin-low subjects, and this human subtype was found to be highly associated with the signature derived from the murine spindloid tumors ($P < 0.0001$) (Fig. S6).

MicroRNA expression profiles are also associated with human breast cancer pathological features and molecular subtypes (31–33). Using a signature of ~50 significantly differentially expressed miRNAs that distinguish claudin-low tumors from other human breast cancer subtypes (33), we identified a near-homogeneous clustering of MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre spindloid tumors that were highly associated with the signature ($P = 0.0004$) (Fig. 3*C* and Dataset S1, Table S8). Notably, consistent with a strong EMT gene expression signature, MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre spindloid tumors showed a significant decrease in expression of miR-200 family members, whose targets include the transcription factors *Zeb1/2* and are known inhibitors of EMT and stemness (34–36). Together, these analyses indicate that MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre spindloid tumors, but not adenocarcinomas, share multiple features in common with human claudin-low breast cancers.

Identification of a Core Claudin-Low Gene Signature. The human claudin-low gene signature constitutes 777 genes (6). To establish whether a restricted, core claudin-low signature could be identified and whether MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre spindloid tumors share common features with human claudin-low tumors, we compared genes systematically highly expressed in MMTV-Met^{mt};Trp53fl/+;Cre spindloid tumors, Trp53fl/+;Cre spindloid tumors, human claudin-low tumors, and human basal B breast cancer cell lines (Fig. 3*D*). This analysis highlighted more than 700 genes that are expressed at elevated levels in either just MMTV-Met^{mt};Trp53fl/+;Cre or just Trp53fl/+;Cre tumors, but not the other, a proportional difference that was significantly higher than expected ($P = 0.009$). When overall gene variance was measured, Trp53fl/+;Cre spindloid tumors were significantly more heterogeneous than MMTV-Met^{mt};Trp53fl/+;Cre spindloid tumors ($P < 2.2 \times 10^{-16}$) (Fig. S7). It is possible that the higher degree of homogeneity observed among MMTV-Met;Trp53fl/+;Cre tumors may result from expression of the MMTV-Met transgene at the point of tumor initiation, whereas *Trp53*-null-alone tumors arise as a result of more stochastic tumorigenic



E

Gene Symbol	Fold change values			
	Met;Trp53fl/+;Cre	Trp53fl/+;Cre	Human CL tumors	Basal B cell lines
CAV1	36.5	28.1	2.0	22.2
VIM	22.8	49.9	2.5	17.3
BCAT1	20.1	39.9	1.5	2.5
SEMA3A	19.6	37.0	2.4	1.3
TWIST1	16.0	27.3	1.5	3.2
VEGFC	15.6	12.8	1.6	3.2
EMP3	14.6	16.0	1.6	17.1
TIMP1	12.6	12.4	1.6	3.5
PROCR	10.1	11.4	1.6	4.2
LAMB1	8.1	11.6	1.8	3.3
IL18	8.1	7.1	1.8	1.9
ITGA5	7.4	10.3	1.6	1.9
MSN	7.2	6.7	2.0	9.3
ZEB2	6.7	7.5	2.8	1.2
TUBB6	5.8	4.0	2.1	4.3
RAB32	5.0	6.1	1.6	2.3
FSTL1	5.0	6.4	2.1	5.6
EVI2A	4.9	5.5	2.3	2.9
RAC2	4.9	4.4	3.5	1.8
AP1S2	4.9	5.3	2.6	2.1
AKR1B1	4.9	3.9	2.1	11.6
FSCN1	4.8	5.1	1.2	3.6
COL6A2	4.8	6.2	1.7	3.1
TPST1	4.6	4.2	1.9	1.3
DSE	4.4	5.4	1.7	3.7
STX2	4.4	4.8	1.5	1.5
GNG11	3.9	4.1	1.8	6.5
FLRT2	3.9	7.7	2.3	3.1
CD44	3.9	3.9	1.9	5.0
TGFBI	3.5	3.2	2.5	6.7
YBX1	3.5	3.2	1.5	1.2
HHEX	3.5	4.7	1.9	1.5
SH3BGRL3	3.4	4.0	1.8	2.1
IFI16	3.2	4.3	2.9	7.6
BMP1	3.1	2.9	1.2	1.5
LEPRE1	3.1	3.2	1.5	2.1

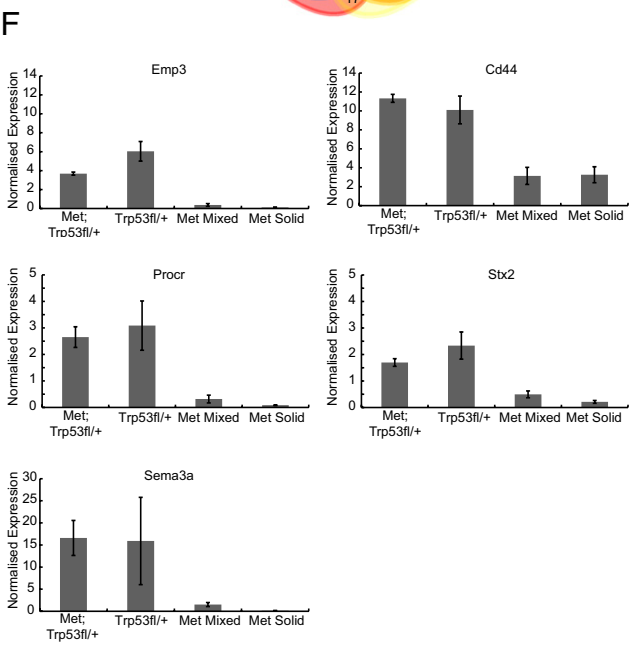


Fig. 3. Gene and miRNA expression profiles of MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre spindloid tumors correlate with those of human claudin-low breast cancer. A cross-species comparison with human breast cancer subtypes reveals that a large proportion of MMTV-Met^{mt};Trp53fl/+;Cre tumors and Trp53fl/+;Cre tumors cluster with the claudin-low molecular subclass at the level of gene expression (*A*). Application of a published claudin-low breast cancer gene expression signature to the mouse model data confirmed this association ($P < 0.0001$) (*B*) and showed that tumors of spindloid pathology were those that correlated with the signature. Similarly, a significant association in miRNA expression was identified through the application of a human claudin-low miRNA signature to MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre tumor data ($P = 4 \times 10^{-4}$) (*C*). To further identify genes associated with claudin-low tumor cell biology and to remove genes expressed by cells in the tumor microenvironment, an intersect of genes highly expressed in human claudin-low breast cancers, MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre spindloid tumors (compared with MMTV-Met^{mt} tumors) and human basal B (claudin-low) breast cancer cell lines, was generated (*D*). This comprised 36 genes (*E*), a selection of which was validated by qRT-PCR (n = 5 MMTV-Met^{mt};Trp53fl/+;Cre, 5 Trp53fl/+;Cre tumors, 3 MMTV-Met^{mt} mixed tumors, and 3 MMTV-Met^{mt} solid tumors), data were normalized to wild-type mammary gland. Error bars, SEM (*F*).

events subsequent to *Trp53* loss. Elevated genes in common between MMTV-Met^{mt};Trp53fl/+;Cre spindloid tumors, human claudin-low tumors, and basal B-cell lines were enriched for signatures related to EMT, HGF signaling, and immune infiltration (Dataset S1, Table S10). In contrast, genes uniquely elevated in Trp53fl/+;Cre spindloid tumors, human claudin-low tumors, and basal B-cell lines (but not MMTV-Met^{mt};Trp53fl/+;Cre spindloid tumors) had enrichment for signatures related to p53 function such as MDM2 and AURKB targets, in addition to apoptosis and chemotherapy response (Dataset S1, Table S10). Hence, although MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre spindloid tumors are more

similar to one another than to MMTV-Met^{mt} tumors (Fig. 2*A*), these tumors are not identical.

In addition to differences, this analysis generated an intersect containing 36 genes in common among MMTV-Met^{mt};Trp53fl/+;Cre spindloid tumors, Trp53fl/+;Cre spindloid tumors, human claudin-low tumors, and human basal B breast cancer cell lines (Fig. 3*D*). Consistent with the highly mesenchymal phenotype of our murine as well as human claudin-low tumors, the core 36-gene intersect includes genes linked to EMT (*Twist1*, *Zeb2*, and *Vim*) in addition to actin cytoskeleton dynamics (*Fscn1*) (37), extracellular matrix interaction, and cell migration (*Msn*, *lamb1*, and *Itga5*) (38, 39) (Fig. 3 *D* and *E*). The 36-gene intersect also included the proinflammatory cytokine *Il-18* and genes associated with poor-outcome breast cancers [*Vegfc* (40) and *Ybx1* (41)]. To test whether the 36-gene intersect alone could identify human claudin-low tumors, we applied it to a human breast cancer dataset containing claudin-low patients (6). Compared with the published claudin-low predictor of Prat et al. (6), which includes 426 genes with elevated expression and 351 genes with decreased, the 36-gene intersect, which represents a small subset, identified claudin-low patients with an equivalent degree of accuracy as the published predictor (Fig. S8) ($P < 0.0001$). Thus, our 36-gene set is functionally equivalent at identifying human claudin-low tumors while elucidating core aspects of claudin-low biology, including potential biomarkers.

Claudin-Low EMT Phenotype Is Dependent on MET Kinase. *Met* was identified within the intersect of MMTV-Met^{mt};Trp53fl/+;Cre tumors, Trp53fl/+;Cre tumors, and basal B-cell lines (Dataset S1, Table S9) and is also retained as part of the published claudin-low predictor (6). To establish whether MET is involved in the maintenance of claudin-low characteristics, primary cells from MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre spindloid tumors, which amplify the endogenous *Met* locus and maintain a strong EMT morphology in culture, were treated with two small-molecule MET-kinase inhibitors (PHA665752 and Crizotinib) (Fig. S9). On inhibition of MET kinase activity, a striking change in cell morphology was observed in both MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre tumor cells. Cells lost their elongated mesenchymal morphology, formed cell-cell junctions positive for the tight junction marker zona occludens protein 1 (ZO-1), and remodeled their actin cytoskeleton with enhanced appearance of cortical actin (Fig. 4*A*). Consistent with the formation of cell-cell junctions and the loss of the EMT morphological phenotype, elevated levels of Claudin 1 protein (CLDN1) were observed (Fig. 4*B*), as well as an elevation in *Cldn1* (Claudin 1) and *Cdh1* (E-cadherin) mRNA (Fig. 4*C*). In contrast, and surprisingly, mRNA levels of EMT transcriptional drivers *Snail*, *Twist*, and *Zeb* were not significantly reduced (Fig. 4*D*). This demonstrates that continued MET signaling has an important role in regulating cell-cell junction disassembly, even in the presence of high levels of key EMT regulators, a characteristic of claudin-low tumor pathology.

In addition to restoring tight junctions and reverting the mesenchymal cell morphology, MET inhibition resulted in significantly impaired proliferation of both MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre spindloid tumor cells, both under normal (adherent) growth conditions and in soft agar (Fig. 5 *A*–*C*). In addition, Annexin V and propidium iodide labeling revealed a significant decrease in the viability of cells that had been treated for 48 h with either PHA665752 or Crizotinib (Fig. 5 *D* and *E*). Together, these data support that both MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre spindloid tumor cells are dependent on MET activity for their proliferation and survival.

MET Inhibition in Vivo Results in Decreased Metastatic Burden. Despite the apparently aggressive phenotype of MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre spindloid tumors, overt lung metastases were not observed. This may be because of the rapid proliferation of the primary tumors, which reach biological endpoint within 2 wk postpalpation. Alternatively, metastasis may be limited by an antitumor immune response, as could be suggested from the gene expression and immune profiling of these tumors (Fig. S5). To establish whether these cells are capable of invasive growth and metastatic spread, as is associated with MET signaling (7), we used a tail vein injection assay to determine whether MMTV-Met^{mt};Trp53fl/+;Cre spindloid tumor cells could grow in the lung microenvironment of immunocompromised mice. Introduction of a firefly luciferase gene allowed visualization of growth in vivo by bioluminescent imaging. MMTV-Met^{mt};Trp53fl/+;Cre spindloid tumor cells were highly aggressive, and by 3 wk postinjection were detected in both the lungs and liver of injected mice, in addition to other sites such as the lymph nodes and peritoneal cavity (Fig. 6). Examination of the lung and liver samples confirmed that MMTV-Met^{mt};Trp53fl/+;Cre tumor cells extravasate and proliferate as lesions external to the blood vessels (Fig. S10), indicating an invasive phenotype. The identification of cells at a variety of anatomical sites in this assay is unusual, as

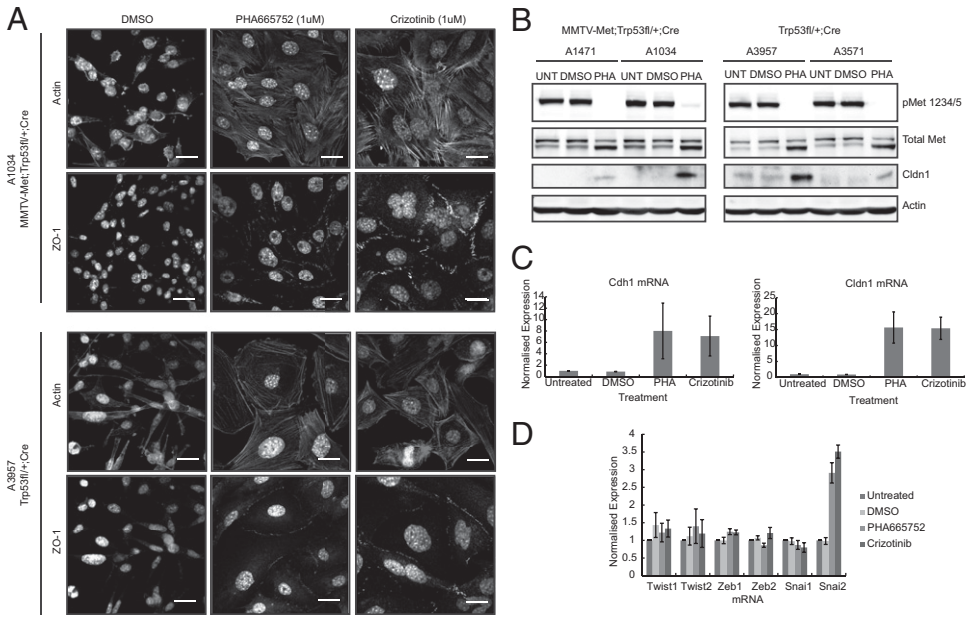
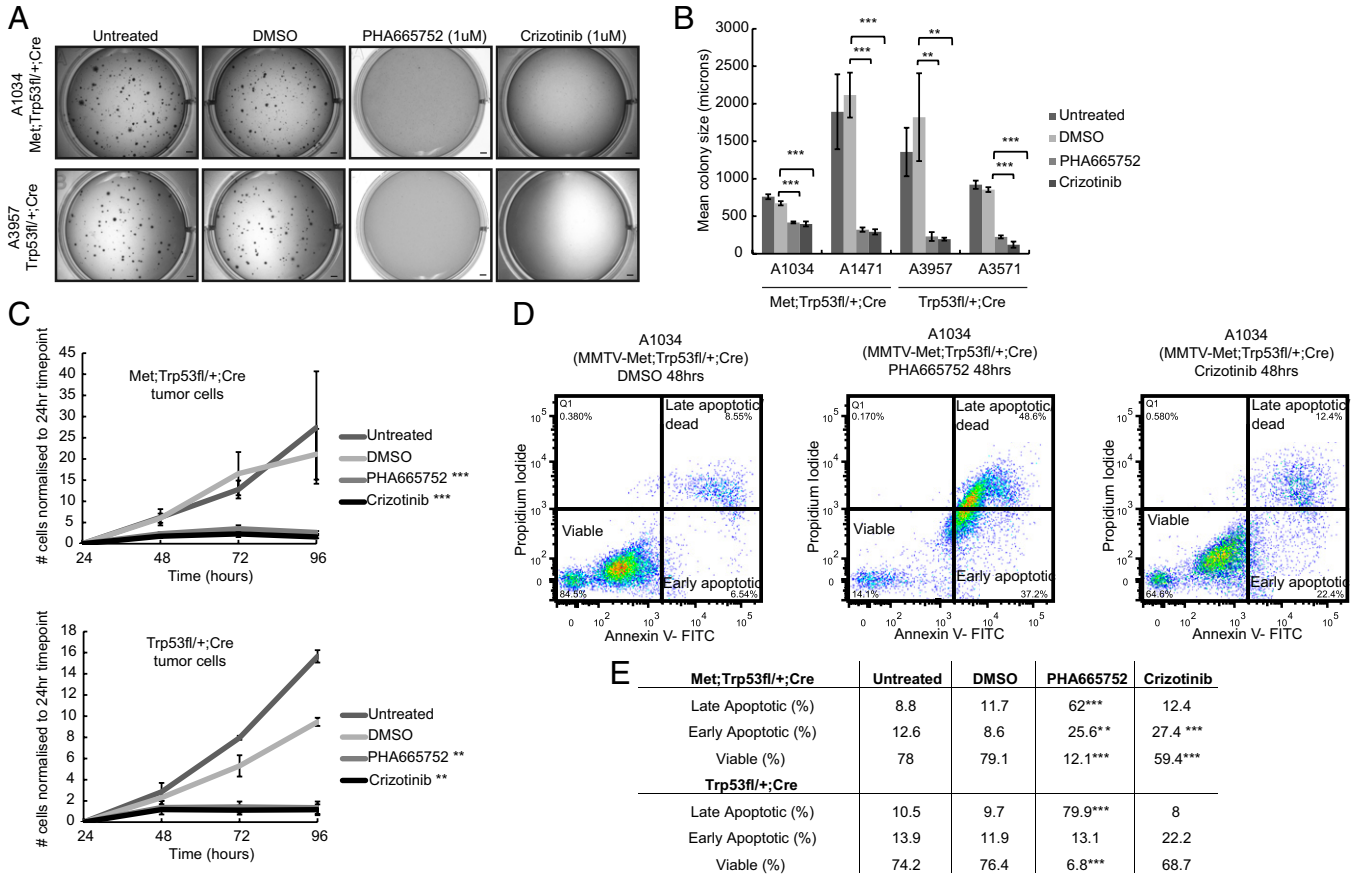


Fig. 4. Treatment of spindloid tumor cells with pharmacological MET inhibitors leads to reversal of the claudin-low phenotype. MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre spindloid tumor cells were treated in vitro with small-molecule inhibitors of MET kinase (PHA665752 [1 μM] or Crizotinib [1 μM]) for 48–72 h. On treatment, cells underwent a distinct morphological change from a mesenchymal to an epithelial-like state (*A*), which included the formation of cell-cell junctions, as demonstrated by the appearance of cortical actin and localization of ZO-1 at sites of cell-cell contact (*A*). (Scale bars, 20 μm.) This was also accompanied by elevated levels of Claudin1 protein, as shown by Western blotting (*B*). Although we also observed an increase in mRNA levels of Claudin1 (*Cldn1*) and E-cadherin (*Cdh1*) on Met inhibition (*C*), there was no corresponding decrease in genes that are well-established as transcriptional drivers of EMT (*Twist1/2*, *Zeb1/2*, and *Snai1/2*) (*D*). Averaged PCR data for four spindloid tumor cell lines (two MMTV-Met^{mt};Trp53fl/+;Cre and two Trp53fl/+;Cre lines) are presented. Error bars, SEM.



E

Met;Trp53fl/+;Cre	Untreated	DMSO	PHA665752	Crizotinib
Late Apoptotic (%)	8.8	11.7	62***	12.4
Early Apoptotic (%)	12.6	8.6	25.6**	27.4 ***
Viable (%)	78	79.1	12.1***	59.4***
Trp53fl/+;Cre				
Late Apoptotic (%)	10.5	9.7	79.9***	8
Early Apoptotic (%)	13.9	11.9	13.1	22.2
Viable (%)	74.2	76.4	6.8***	68.7

Fig. 5. Inactivation of MET kinase inhibits the proliferation and survival of *Met*-amplified spindloid tumor cells. Tumor cells isolated from two MMTV-Met^{mt};Trp53fl/+;Cre and two Trp53fl/+;Cre spindloid mammary tumors formed smaller colonies in soft agar during a 10-d assay in the presence of MET kinase inhibitors (PHA665752 [1 μM] and Crizotinib [1 μM]); representative images for two cell lines are shown (*A*). (Scale bars, 1,000 μm.) Reduction in colony size was highly significant in all four cell lines (*B*). Error bars, SEM. Significantly impaired proliferation resulting from MET inhibition was also demonstrated in a 4-d proliferation assay in which the same cell lines were grown on tissue culture plastic and counted every 24 h (*C*). Error bars, SEM. To assess any effect on cell viability, cells treated with MET inhibitors for 48 h were stained with Annexin-V and propidium iodide and analyzed by flow cytometry. Representative plots for one MMTV-Met^{mt};Trp53fl/+;Cre cell line are shown (*D*), and averaged data for two MMTV-Met^{mt};Trp53fl/+;Cre and two Trp53fl/+ cell lines are tabulated (*E*). All four cell lines responded similarly and showed a dramatic increase in the proportion of cells in late-stage apoptosis after treatment with PHA665752 (e.g., 11.7% of MMTV-Met^{mt};Trp53fl/+;Cre cells were in late apoptosis in the DMSO control vs. 62% in the PHA667572 treatment). The effect of Crizotinib on cell viability was more moderate (only 12.4% of MMTV-Met^{mt};Trp53fl/+;Cre cells treated with Crizotinib were in late apoptosis). ***$P < 0.01$; **$P < 0.05$ (*E*).

cells introduced via the tail vein bypass the normal metastatic cascade and are delivered directly to the lung, only rarely being detected in other organs (42–44). Notably, daily treatment of injected mice with the orally available MET inhibitor Crizotinib (45 mg·kg^{-1}·d^{-1}) significantly reduced metastatic growth both in the lungs and livers of the mice (Fig. 6), showing that the metastatic growth of these EMT mammary tumor cells is highly dependent on MET activity.

Elevated MET and TP53 Protein Correlates with Hormone Receptor-Negative Status and Poor Prognosis in Human Breast Cancer. Alterations in *TP53* are typically associated with the basal subtype of TN breast cancer (2). Missense mutations are associated with increased stability of the TP53 protein and can be detected by IHC analysis, as significantly higher tumor tissue staining is observed compared with tumors with *TP53* truncating mutations or wild-type *TP53* (45). Overexpression of MET and expression of mutant TP53 proteins have both been shown to have prognostic value individually; however, the significance of their coexistence in the same tumor has not been examined. The examination of MET and TP53 protein in a cohort of 618 axillary lymph node–negative human breast cancer cases (46) revealed that tumor epithelium was positive for MET immunostaining and/or TP53 staining, with an absence of staining in the stroma (Fig. 7*A*). Tumors that stained strongly for MET were more likely to be TP53 positive than those negative for MET, as 13.9% of all 618 tumors studied were MET+/TP53+ compared with 9.1% that were MET−/TP53+ (Fig. 7*B*) ($P < 0.0001$).

Tumors that scored for both high MET and TP53 were observed in all histological subtypes, but a significantly greater proportion of MET/TP53 copositive tumors were estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and CK5-positive (61%, 71%, and 44%, respectively) than tumors with other combinations of MET and TP53 (24%, 38%, and 14%, respectively; $P < 0.0001$) (Dataset S1, Table S11). Basal, TN phenotype (TNP)-nonbasal, Her2, and luminal subtypes were determined as previously described (47). MET/P53 copositive tumors were found to correlate most significantly with the basal ($P < 0.0001$) and TNP-nonbasal ($P = 0.0211$) subtypes (Table 2). More precise identification of claudin-low patients would require an examination of a claudin-low gene expression signature within this set and/or the use of a positive IHC marker for claudin-low, which is currently not known. However, on the basis of the available information for this cohort, both of these subtypes could include patients of claudin-low pathology.

The majority of MET/TP53-positive tumors (94%) scored high for cell proliferation marker KI67 compared with 57% for other combinations of MET and TP53 ($P < 0.0001$) (Dataset S1, Table S11). Consistent with this, combined MET/TP53-positive tumor status correlates with poor disease-free survival among lymph node–negative patients (Fig. 7*C*; log rank $P = 0.0012$) compared

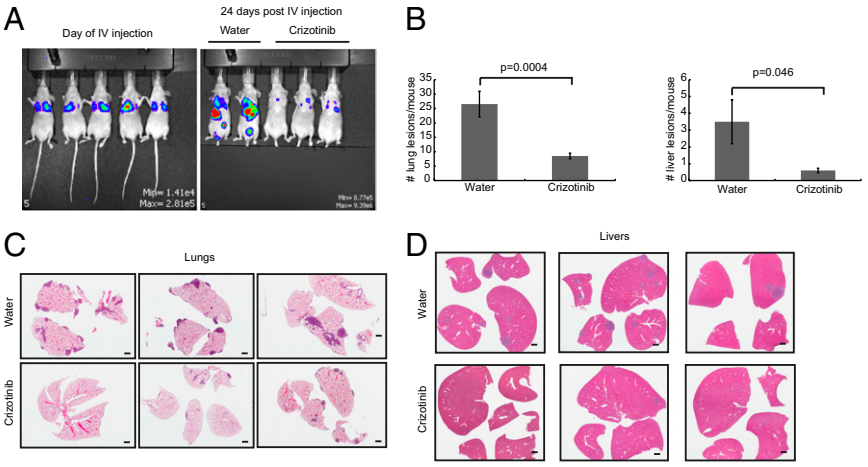


Fig. 6. MET inhibition impairs the metastatic potential of spindloid mammary tumor cells. An MMTV-Met^{mt};Trp53fl/+;Cre spindloid tumor cell line expressing firefly luciferase was injected i.v. by the tail vein into 35 nude mice (0.5×10^6 cells/mouse). Mice were imaged on the day of injection (*A*) and twice per week thereafter to monitor the development of metastases. A control group of 15 mice was gavaged daily with water and compared with 20 mice receiving a daily gavage of Crizotinib (45 mg/kg/d). By day 24, control mice showed extensive metastatic burden compared with Crizotinib-treated mice (*A*). Lungs and livers were harvested from all animals at day 24 and scored histologically for metastatic lesions. Mice treated with Crizotinib showed a significant reduction in the number of lesions detected in both the lungs and liver (*B*). Representative histology from three control and three Crizotinib-treated mice is shown (*C* and *D*).

with patients with other combinations of MET/TP53 status, demonstrating that the combination of elevated MET with positive TP53 IHC is a strong predictor of poor outcome. This association persisted in multivariate analysis after adjustment for traditional histopathological prognostic factors (Dataset S1, Tables S11 and S12). Finally, MET/TP53 copositivity can also identify poor-outcome patients within the TN group alone (Fig. 7*D*). Together, these results strongly support a role for MET/TP53 signaling in human ER/PR-negative, CK5-positive breast cancers and in breast cancers with high KI67 staining and poor outcome.

Discussion

One of the challenges for the effective treatment of breast cancer is the heterogeneity of the disease (48). TN breast cancers alone encompass at least 2 (and potentially 6, some of which are more recently identified) (49) molecular subtypes referred to as basal-like and claudin-low (3, 6), for which there are a lack of known therapeutic targets and suitable animal models. Evidence supports that the MET RTK is elevated in human TN breast cancers (8). This, together with the observation that murine models expressing a weakly activated Met in the mammary epithelium develop tumors with basal-like characteristics, supports a role for MET in the development of basal-like mammary tumors (14, 15). However, the involvement of MET in other subtypes within TN or the ability of MET to synergize with known alterations in TN breast cancer has not been addressed. To create a more accurate model for human TN breast cancer, we have exploited the frequent occurrence of *TP53* mutations in TN breast cancer and generated a model combining expression of a weakly oncogenic MET receptor (MMTV-Met^{mt}) (14) with conditional deletion of *Trp53* in the mammary glands of FVB/N mice (MMTV-Met^{mt};Trp53fl/+;MMTV-Cre-recombinase). The resulting MMTV-Met^{mt};Trp53fl/+;Cre mouse model shows effective cooperation of *Met* with *Trp53* loss in mammary tumorigenesis, manifested as a significant increase in tumor penetrance over both MMTV-Met^{mt} and Trp53fl/+;Cre control groups.

Notably, the majority of mammary tumors that form in the MMTV-Met^{mt};Trp53fl/+;Cre model (80%) share molecular features and histological markers of the claudin-low subtype of human TN breast cancer (6). Key aspects include enrichment for a claudin-low gene expression signature ($P < 0.0001$) (6) and miRNA signature, including loss of Claudin gene expression (e.g., *Cldn1*, *Cldn3*, *Cldn4*, and *Cldn7*), expression of the core EMT gene signature (*Snai1/2*, *Twist1/2*, and *Zeb1/2*), and lymphocytic infiltration (6, 23). This phenotype is shared by 5/8 Trp53fl/+;Cre tumors, which, in addition to loss of *Trp53*, show amplification of *Met* and a similar claudin-low gene expression signature to MMTV-Met^{mt};Trp53fl/+;Cre spindloid tumors. In contrast, MMTV-Met^{mt} tumors clustered with basal and luminal subtypes (14), and only a single MMTV-Met^{mt} tumor with a spontaneous *Trp53* mutation, expressed a claudin-low signature (Fig. 3*B*). One important difference within Trp53fl/+;Cre tumors is that Met amplification was not detected in Trp53fl/+;Cre tumors of adenocarcinoma pathology. This indicates that loss of *Trp53* alone, as evident in Trp53fl/+;Cre adenocarcinomas, is insufficient for spindloid pathology and a penetrant claudin-low phenotype and supports a synergistic role for *Met*, together with *Trp53* loss, in promoting tumors with a spindloid pathology and claudin-low molecular subtype in the FVB background. This is consistent with the enhanced penetrance (70%) and high incidence of spindloid (80%), claudin-low-type tumors in MMTV-Met^{mt};Trp53fl/+;Cre mice.

Compared with other mouse mammary tumor models, MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre tumors of spindloid pathology clustered together and in close proximity to tumors belonging to models such as p53-null transplants, in addition to 7,12-Dimethylbenz(a)anthracene (DMBA), MMTV-CreBrca1$^{co/co}$, and whey acidic protein (WAP)-Myc (Fig. S11). Interestingly, the WAP-Myc model can also induce tumors of spindloid pathology (3), and amplification of the *Myc* locus is observed in 3 of 7 of the MMTV-Met^{mt};Trp53fl/+;Cre spindloid tumors and 47% of human claudin-low tumors (23). However,

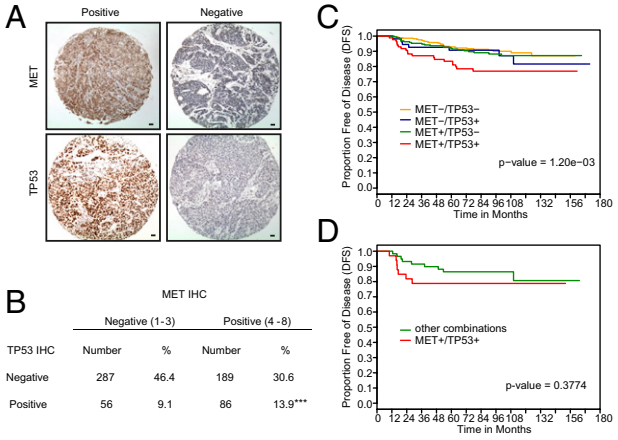


Fig. 7. Elevated MET expression in human breast cancer is associated with TP53 mutation and combining MET with TP53 positive IHC identifies patients with poor prognosis. A human breast cancer tissue microarray comprising 618 node-negative patients was stained for MET and TP53 (*A*). Analysis showed that MET-positive tumors were more likely to stain positively for TP53 (indicative of mutated TP53) than MET-negative tumors (*B*) and that patients with MET-positive–TP53-positive tumors had a significantly worse outcome than patients with either MET or TP53 positivity alone ($P = 0.0012$) (*C*). Within TN patients specifically ($n = 93$), there was a trend toward MET-TP53 copositivity correlating with a poorer outcome ($P = 0.3774$), with a clear separation from patients with other combinations of MET and TP53 IHC within the first 36 mo after diagnosis.

Table 2. Association of MET-positive-, TP53-positive breast tumors with the basal and TNP-nonbasal subtypes

	MET+/TP53+ (n = 86)		Other combinations of MET and TP53 (n = 532)		
Subgroup	No.	%	No.	%	P
Basal					
Yes	26	**30.2**	42	**7.9**	<0.0001
No	60	69.8	490	92.1	
TNP-nonbasal					
Yes	6	**7.0**	11	**2.0**	0.0211
No	80	93.0	521	98.0	

Scoring for MET and TP53 IHC on a human breast cancer tissue microarray was correlated with subtype. Breast cancers that stained positively for both MET and TP53 were more likely to be classified as basal, than breast cancers with other combinations of MET and TP53 staining (30.2% vs. 7.9%). Likewise, more MET/TP53 copositive breast cancers were classified as TNP-nonbasal, than breast cancers positive to MET or TP53 alone (7.0% vs. 2.0%).

although 80% of the MMTV-Met^{mt};Trp53fl/+;Cre tumors described here are spindloid or contain a spindle-cell component, only a fraction of tumors in the aforementioned models display this phenotype (3). Hence, MMTV-Met^{mt};Trp53fl/+;Cre tumors represent a robust model for efficient induction of claudin-low breast cancer. Similarly, only 10% of tumors arising in a transplant model of *Trp53*-null mammary epithelium display a claudin-low phenotype (50), providing further evidence that loss of *Trp53* may be insufficient for this phenotype. Consistent with this, all Trp53fl/+;Cre tumors of spindloid pathology, correlating with a claudin-low subtype, contained amplification of the *Met* locus and variable adjacent genes. This links MET and P53 synergistically in promoting spindloid pathology and claudin-low like tumors in the FVB genetic background, especially as Trp53fl/+;Cre tumors of adenocarcinoma pathology did not amplify *Met* (Fig. S3*A*).

The mechanism selecting for *Met* amplification in the Trp53fl/+;Cre FVB model is unclear. A similar amplification of Met is observed in 73% of mammary tumors involving germ-line loss of *Trp53* in combination with a conditional breast cancer 1 (*Brca1*) mutation (Brca1$^{\Delta 11/co}$;MMTV-Cre;Trp53$^{+/-}$) (16). However, although *Met* amplification in cell lines established from Brca1$^{\Delta 11/co}$;MMTV-Cre;Trp53$^{+/-}$ tumors was carried on double minutes and lost from cells in culture (16), *Met* amplification in cell lines derived from MMTV-Met;Trp53fl/+;Cre and Trp53fl/+;Cre tumors is stable and retained during serial passage (Fig. S12). Moreover, these cell lines are continuously dependent on MET signaling for their EMT phenotype, as well as for their proliferation and survival both in culture and in vivo. Thus, *Met* amplification with consequent constitutive activation of the kinase is required to maintain the claudin-low mesenchymal phenotype of these cells. The unstable nature of the *Met* amplicon in the Brca1$^{\Delta 11/co}$;MMTV-Cre;Trp53$^{+/-}$ model may reflect loss of function of *Brca1,* which contributes to chromosomal instability, whereas we observe no decrease in *Brca1* or *Brca2* expression in MMTV-Met;Trp53fl/+;Cre tumors compared with normal mammary gland (Dataset S1, Table S2). Interestingly, an amplicon containing *Met* was also recently detected in murine mammary tumors that arise as a result of potentiated Notch signaling and that also model both basal-like and claudin-low breast cancers (51). Although the stability of this amplicon was not addressed in this study, this lends further support for a specific role for MET signaling in murine models of claudin-low breast cancer.

Cell explants derived from MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre spindloid claudin-low-like tumors retain a mesenchymal phenotype that is highly dependent on continued MET signaling. When treated with two pharmacological MET inhibitors, a reversal of the EMT morphological phenotype was observed, with elevated levels of Claudin 1 and reformation of ZO-1 positive cell-cell junctions, which are claudin-dependent (52). Although the effect of MET signaling on tight junction disassembly is clear, we observed no changes in the mRNA levels of the core transcriptional drivers of EMT (*Snai1/2*, *Twist1/2*, and *Zeb1/2*) on MET inhibition (Fig. 4), demonstrating that continued MET activation is essential to maintain the EMT morphological phenotype and the loss of claudin gene expression, a hallmark of human claudin-low tumors (6).

Although MET can promote elevated expression of *Zeb1* and *Snail* to initiate EMT (14), the core EMT signature is elevated in Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre spindloid tumors compared with MMTV-Met^{mt} basal subtype tumors. This likely reflects the role for wild-type *Trp53* in promoting an epithelial phenotype through transcriptional activation of the miR-200 family (underexpressed within the human claudin-low miRNA signature) that negatively regulates the key regulators of EMT (34). Consistent with this, after loss of *Trp53* in MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre tumors, we observe a decrease in the miR-200 family and correspondingly high levels of EMT transcriptional drivers that are not altered after MET inhibition.

Accumulating evidence supports a role for MET and MET-dependent signals in human claudin-low breast cancer. MET contributes to a published claudin-low predictor (6). A strong MET signaling network is present in both MMTV-Met^{mt};Trp53fl/+;Cre and Trp53l/+;Cre tumors [*Hgf*, *Cd44*, *Plaur* (plasminogen activator, urokinase receptor), *Plau* (plasminogen activator, urokinase), *Ets1* and *Ybx1*] (28–30, 53, 54), elements of which are also represented in the 36-gene intersect formed with human claudin-low tumors and basal B-cell lines (*Cd44* and *Ybx1*) (Fig. 3*E*). The selection for amplification of the *Met* locus in *Trp53*- null tumors of spindloid pathology is striking and highlights an emerging concept in cancer whereby genes that function synergistically to enhance signaling will frequently be coselected during tumor formation or progression.

We propose that *Met* synergizes in this context with loss of function of *Trp53* but may also synergize with other regulators of this phenotype such as Notch (51). The observed amplification of genes also amplified in human basal and claudin-low breast cancer such as *Caveolin 1* and *Myc* in the MMTV-Met^{mt};Trp53fl/+;Cre model provides a valuable tool to understand the molecular events and signaling pathways that drive TN breast cancers. This model also presents an opportunity to study the tumor microenvironment of claudin-low breast cancer, as demonstrated by the evidence for robust leukocyte infiltration. Because human claudin-low breast cancer is especially difficult to treat due to the lack of biomarkers, determining molecular targets that can be used in drug therapy is of utmost importance. In addition, because small-molecule MET inhibitors are presently in clinical trials for multiple cancers, this raises the possibility that TP53 status may be important for patient selection.

Materials and Methods

Transgenic Mice. MMTV-Met^{mt} mice were described previously (14). MMTV-Cre mice were generated in the laboratory of W.J. Muller (55). Mice with floxed-*Trp53* alleles are described elsewhere (21), were obtained from the National Cancer Institute mouse repository, and were bred onto a pure FVB background. Mice were housed in accordance with McGill University Animal Ethics Committee guidelines.

Immunohistochemical and Immunofluorescent Analyses of Mouse Tissue and Cell Lines. Cells were fixed and histology samples prepared as described in *SI Materials and Methods*. Primary and secondary antibodies are detailed in Dataset S1, Table S13.

Microarray Data. Gene expression profiles were generated using Agilent 4 × 44K whole-mouse genome gene expression microarrays. Copy number gains and losses were assessed using Agilent 4 × 44K whole-mouse genome CGH arrays. miRNA profiling was performed using the Agilent 8 × 15K mouse miRNA platform. Raw and normalized microarray data have been deposited in the Gene Expression Omnibus database under accession no. GSE41748. All analyses are detailed in the *SI Materials and Methods*.

Isolation and Culture of Mouse Mammary Tumor Cells. Primary cells were isolated from mouse mammary tumors as described (56). Cells were cultured

in DMEM supplemented with 5% (vol/vol) serum, epidermal growth factor (5 ng/mL), insulin (5 μg/mL), bovine pituitary extract (35 μg/mL), and hydrocortisone (1 μg/mL).

Met Inhibition. MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre tumor cell lines were treated with PHA665752 (Pfizer) or Crizotinib (LC Laboratories) at a final concentration of 1 μM. Control cells were incubated with an equivalent concentration of DMSO alone for the same amount of time.

ACKNOWLEDGMENTS. We thank Anie Monast for maintenance of our transgenic mouse colony and genotyping. We also thank Dr. Peter Siegel and Dr. Josie Ursini-Siegel for helpful discussions, the Goodman Cancer Research Centre Histology service, Ken McDonald and Diane Ethier for assistance with flow cytometry, and Dr. Pierre Lepage at the Genome Quebec Innovation Centre for loss-of-heterozygosity analysis. The Met inhibitor PHA665752 was a kind gift from Pfizer. We acknowledge infrastructure support and technical assistance from the Breast Cancer Functional Genomics Group, which is partially supported by funds from the Terry Fox New Frontiers Program. This research was supported with funds from a Terry Fox New Frontier grant (to M.P). in addition to funds from the Canadian Institutes for Health Research (to I.L.A. and M.P.)., a Susan G. Komen for the Cure postdoctoral fellowship (to J.F.K.), a McGill Faculty of Medicine Studentship (to S.M.S.), and a Canadian Institutes for Health Research studentship (to R.L.).

1. Ferlay et al. (2008) Globocan 2008 v2.0, Cancer Incidence and Mortality Worldwide. *IARC CancerBase* 10. Available at https://globocan.iarc.fr.
2. Sørlie T, et al. (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA* 98(19):10869–10874.
3. Herschkowitz JI, et al. (2007) Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol* 8(5):R76.
4. Brenton JD, Carey LA, Ahmed AA, Caldas C (2005) Molecular classification and molecular forecasting of breast cancer: Ready for clinical application? *J Clin Oncol* 23(29):7350–7360.
5. Sorlie T, et al. (2003) Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci USA* 100(14):8418–8423.
6. Prat A, et al. (2010) Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res* 12(5):R68.
7. Birchmeier C, Birchmeier W, Gherardi E, Vande Woude GF (2003) Met, metastasis, motility and more. *Nat Rev Mol Cell Biol* 4(12):915–925.
8. Ponzo MG, Park M (2010) The Met receptor tyrosine kinase and basal breast cancer. *Cell Cycle* 9(6):1043–1050.
9. Camp RL, Rimm EB, Rimm DL (1999) Met expression is associated with poor outcome in patients with axillary lymph node negative breast carcinoma. *Cancer* 86(11): 2259–2265.
10. Garcia S, et al. (2007) Poor prognosis in breast carcinomas correlates with increased expression of targetable CD146 and c-Met and with proteomic basal-like phenotype. *Hum Pathol* 38(6):830–841.
11. Ghoussoub RA, et al. (1998) Expression of c-met is a strong independent prognostic factor in breast carcinoma. *Cancer* 82(8):1513–1520.
12. Taniguchi T, et al. (1995) Serum concentrations of hepatocyte growth factor in breast cancer patients. *Clin Cancer Res* 1(9):1031–1034.
13. Tyan SW, et al. (2011) Breast cancer cells induce cancer-associated fibroblasts to secrete hepatocyte growth factor to enhance breast tumorigenesis. *PLoS ONE* 6(1):e15313.
14. Ponzo MG, et al. (2009) Met induces mammary tumors with diverse histologies and is associated with poor outcome and human basal breast cancer. *Proc Natl Acad Sci USA* 106(31):12903–12908.
15. Graveel CR, et al. (2009) Met induces diverse mammary carcinomas in mice and is associated with human basal breast cancer. *Proc Natl Acad Sci USA* 106(31): 12909–12914.
16. Smolen GA, et al. (2006) Frequent met oncogene amplification in a Brca1/Trp53 mouse model of mammary tumorigenesis. *Cancer Res* 66(7):3452–3455.
17. He L, et al. (2007) A microRNA component of the p53 tumour suppressor network. *Nature* 447(7148):1130–1134.
18. Perou CM, et al. (2000) Molecular portraits of human breast tumours. *Nature* 406(6797):747–752.
19. Lim E, et al.; kConFab (2009) Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat Med* 15(8): 907–913.
20. Cardiff RD (2010) The pathology of EMT in mouse mammary tumorigenesis. *J Mammary Gland Biol Neoplasia* 15(2):225–233.
21. Jonkers J, et al. (2001) Synergistic tumor suppressor activity of BRCA2 and p53 in a conditional mouse model for breast cancer. *Nat Genet* 29(4):418–425.
22. Junttila MR, Evan GI (2009) p53—a Jack of all trades but master of none. *Nat Rev Cancer* 9(11):821–829.
23. Weigman VJ, et al. (2012) Basal-like Breast cancer DNA copy number losses identify genes involved in genomic instability, response to therapy, and patient survival. *Breast Cancer Res Treat* 133(3):865–880.
24. White DE, Cardiff RD, Dedhar S, Muller WJ (2001) Mammary epithelial-specific expression of the integrin-linked kinase (ILK) results in the induction of mammary gland hyperplasias and tumors in transgenic mice. *Oncogene* 20(48):7064–7072.
25. Taube JH, et al. (2010) Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. *Proc Natl Acad Sci USA* 107(35):15449–15454.
26. Kitamura T, Taketo MM (2007) Keeping out the bad guys: Gateway to cellular target therapy. *Cancer Res* 67(21):10099–10102.
27. Groom JR, Luster AD (2011) CXCR3 ligands: Redundant, collaborative and antagonistic functions. *Immunol Cell Biol* 89(2):207–215.
28. Matzke A, et al. (2007) Haploinsufficiency of c-Met in cd44-/- mice identifies a collaboration of CD44 and c-Met in vivo. *Mol Cell Biol* 27(24):8797–8806.
29. Gambarotta G, et al. (1996) Ets up-regulates MET transcription. *Oncogene* 13(9): 1911–1917.
30. Finkbeiner MR, et al. (2009) Profiling YB-1 target genes uncovers a new mechanism for MET receptor regulation in normal and malignant human mammary cells. *Oncogene* 28(11):1421–1431.
31. Blenkiron C, et al. (2007) MicroRNA expression profiling of human breast cancer identifies new markers of tumor subtype. *Genome Biol* 8(10):R214.
32. Iorio MV, et al. (2005) MicroRNA gene expression deregulation in human breast cancer. *Cancer Res* 65(16):7065–7070.
33. Buffa FM, et al. (2011) microRNA-associated progression pathways and potential therapeutic targets identified by integrated mRNA and microRNA expression profiling in breast cancer. *Cancer Res* 71(17):5635–5645.
34. Chang CJ, et al. (2011) p53 regulates epithelial-mesenchymal transition and stem cell properties through modulating miRNAs. *Nat Cell Biol* 13(3):317–323.
35. Kim T, et al. (2011) p53 regulates epithelial-mesenchymal transition through microRNAs targeting ZEB1 and ZEB2. *J Exp Med* 208(5):875–883.
36. Gregory PA, et al. (2008) The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nat Cell Biol* 10(5):593–601.
37. Sedeh RS, et al. (2010) Structure, evolutionary conservation, and conformational dynamics of Homo sapiens fascin-1, an F-actin crosslinking protein. *J Mol Biol* 400(3):589–604.
38. Nürnberg A, Kitzing T, Grosse R (2011) Nucleating actin for invasion. *Nat Rev Cancer* 11(3):177–187.
39. Valastyan S, Benaich N, Chang A, Reinhardt F, Weinberg RA (2009) Concomitant suppression of three target genes can explain the impact of a microRNA on metastasis. *Genes Dev* 23(22):2592–2597.
40. Mohammed RA, et al. (2007) Prognostic significance of vascular endothelial cell growth factors -A, -C and -D in breast cancer and their relationship with angio- and lymphangiogenesis. *Br J Cancer* 96(7):1092–1100.
41. Stratford AL, et al. (2007) Epidermal growth factor receptor (EGFR) is transcriptionally induced by the Y-box binding protein-1 (YB-1) and can be inhibited with Iressa in basal-like breast cancer, providing a potential target for therapy. *Breast Cancer Res* 9(5):R61.
42. Goodale D, Phay C, Postenka CO, Keeney M, Allan AL (2009) Characterization of tumor cell dissemination patterns in preclinical models of cancer metastasis using flow cytometry and laser scanning cytometry. *Cytometry A* 75(4):344–355.
43. Francia G, Cruz-Munoz W, Man S, Xu P, Kerbel RS (2011) Mouse models of advanced spontaneous metastasis for experimental therapeutics. *Nat Rev Cancer* 11(2):135–141.
44. Du YC, Chou CK, Klimstra DS, Varmus H (2011) Receptor for hyaluronan-mediated motility isoform B promotes liver metastasis in a mouse model of multistep tumorigenesis and a tail vein assay for metastasis. *Proc Natl Acad Sci USA* 108(40): 16753–16758.
45. Ozcelik H, Pinnaduwage D, Bull SB, Andrulis IL (2007) Type of TP53 mutation and ERBB2 amplification affects survival in node-negative breast cancer. *Breast Cancer Res Treat* 105(3):255–265.
46. Mulligan AM, Pinnaduwage D, Bull SB, O'Malley FP, Andrulis IL (2008) Prognostic effect of basal-like breast cancers is time dependent: Evidence from tissue microarray studies on a lymph node-negative cohort. *Clin Cancer Res* 14(13):4168–4174.
47. Voduc KD, et al. (2010) Breast cancer subtypes and the risk of local and regional relapse. *J Clin Oncol* 28(10):1684–1691.
48. Curtis C, et al.; METABRIC Group (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 486(7403):346–352.
49. Lehmann BD, et al. (2011) Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest* 121(7):2750–2767.
50. Herschkowitz JI, et al. (2012) Comparative oncogenomics identifies breast tumors enriched in functional tumor-initiating cells. *Proc Natl Acad Sci USA* 109(8):2778–2783.
51. Xu K, et al. (2012) Lunatic fringe deficiency cooperates with the Met/Caveolin gene amplicon to induce basal-like breast cancer. *Cancer Cell* 21(5):626–641.
52. Furuse M (2010) Molecular basis of the core structure of tight junctions. *Cold Spring Harb Perspect Biol* 2(1):a002907.
53. Zöller M (2011) CD44: Can a cancer-initiating cell profit from an abundantly expressed molecule? *Nat Rev Cancer* 11(4):254–267.
54. Jeffers M, Rong S, Vande Woude GF (1996) Enhanced tumorigenicity and invasion-metastasis by hepatocyte growth factor/scatter factor-met signalling in human cells concomitant with induction of the urokinase proteolysis network. *Mol Cell Biol* 16(3): 1115–1125.
55. Andrechek ER, et al. (2000) Amplification of the neu/erbB-2 oncogene in a mouse model of mammary tumorigenesis. *Proc Natl Acad Sci USA* 97(7):3444–3449.
56. Ling C, Zuo D, Xue B, Muthuswamy S, Muller WJ (2010) A novel role for 14-3-3sigma in regulating epithelial cell polarity. *Genes Dev* 24(9):947–956.

Supporting Information

Knight et al. 10.1073/pnas.1210353110

SI Materials and Methods

Immunohistochemical and Immunofluorescent Analyses of Mouse Tissue and Cell Lines. Histology samples were fixed for 24 h in 10% formalin, embedded in paraffin, and sectioned at 5 μm. Sections were stained with H&E and reviewed by an experienced comparative pathologist (R.D.C.).

Antigen retrieval of deparaffinized tissue sections was performed in boiling 10 mM citrate buffer at pH 6.0 for most antigens, or 10 mM Tris-base/1 mM EDTA solution as indicated in Dataset S1, Table S11. F4/80 staining was performed on frozen sections. Tissue sections were blocked for 10 min with Universal Blocking Agent (Biogenics). Primary and secondary antibodies were diluted in 2% BSA in PBS and are detailed in Dataset S1, Table S13. Immunohistochemical labeling was detected using the Vectastain Elite ABC kit (Vector Laboratories) and 3–3′-diaminobenzidine.

For immunofluorescent labeling of cell lines, cells were cultured on glass coverslips. Cells were fixed for 10 min in 2% paraformaldehyde at room temperature. Primary and secondary antibodies were used as indicated in Dataset S1, Table S13.

Microscopy and Imaging. Phase contrast microscopy was performed using an Olympus CKX41 microscope, and images were taken using a Lumenera Infinity 1 digital camera.

Stained tissue sections were imaged using an Aperio-XT slide scanner (Aperio Technologies).

For immunofluorescence, fluorophore-conjugated secondary antibodies are listed in Dataset S1, Table S13. Images were taken using an LSM510 confocal microscope (Carl Zeiss) and analyzed using Zen software.

Gene Expression Microarray Data. RNA was extracted from mouse mammary tumors and normal mammary glands that had been snap-frozen immediately after animal necropsy. Tissues were powdered under liquid nitrogen and homogenized in Qiashredder columns, and RNA was isolated using the Qiagen Allprep kit. The quality of the RNA was checked using a Bioanalyser (Agilent), and quantifications were made using a Nanodrop (Thermo Fisher).

One round of amplification and labeling for microarray hybridization was carried out using the Amino Allyl MessageAmp II aRNA kit (Ambion AM1753). Universal Mouse reference RNA (Stratagene catalog no. 740100–41) was amplified and labeled in the same manner.

Next, 825 ng of Cy3-labeled aRNA samples were cohybridized with 825 ng of Cy5-labeled reference aRNA to whole-mouse genome (4 × 44K) arrays (Agilent, G4122F). Slides were washed according to the manufacturer's instructions and scanned using an Agilent dual-laser scanner (G2505B). Feature extraction was performed using Agilent software (FE 9.5.3.1).

Array data were normalized as in ref. 1, and analyses were carried out in the R statistical framework with Bioconductor. All hierarchical clustering used Ward's agglomeration algorithm with an Euclidean distance metric. Unsupervised class discovery was performed by filtering to include only probes with an interquartile range of at least 2 across all samples. Mouse–human orthologs were determined using the biomaRt package (2).

Comparisons with other datasets were made by first separately column- and row-scaling genes in each dataset to ~N(0,1) and then combining the datasets over a filtered set of genes representing the cross-species or murine intrinsic gene lists derived in ref. 3. Human tumor subtype classifications were the same used as those in ref. 3, which used an unsupervised clustering approach over a set of highly variable probes. Differentially expressed genes were identified using limma (4) with the Benjamini-Hochberg method to adjust for multiple testing (5). To further reduce the number of genes identified in our murine samples, probes were additionally required to have at least a 1.5 log 2-fold change in determining the 36-gene intersect between human tumors and cell lines. When applying a signature to a dataset, samples were either hierarchically clustered or ordered by a modified rank-sum of their genes. That is, signature genes expected to have elevated expression were ranked in ascending order, whereas genes expected to have decreased were ranked in descending order across all tumors. These ranks were summed for each sample and then normalized to the number of nonmissing values for that sample. A final tumor ordering was made by ranking all of these normalized sums from least to greatest.

The significance of an association between a signature and a given subgroup was determined using Gene Set Enrichment Analysis (6, 7) with 10,000 sample permutations. *P* values for the up and down lists of the same signature were combined using Fisher's method. Enrichment of our gene sets for previously published signatures was determined using a hypergeometric test, followed by Benjamini-Hochberg adjustment for multiple testing. The ~6,500 signatures tested were an amalgamation primarily of those obtained from The Molecular Signatures Database (MSigDB) (5), GenesigDB (8), and various other signatures collected from the literature. The expected proportion of differentially expressed genes shared between MMTV-Metmt;Trp53fl/+;MMTV-Cre recombinase and Trp53fl/+;MMTV-Cre recombinase tumors was determined using 10,000 sample permutations. The heterogeneity of tumor types was determined by measuring the variance across samples for all genes on the array, and statistical differences in these distributions were determined using a one-sided Kolmogorov-Smirnov test.

Array-Comparative Genomic Hybridization Data. Genomic DNA was isolated from snap-frozen tissue pieces using the Qiagen Allprep kit (as described for RNA isolation). DNA was prepared for array hybridization using the Agilent Genomic DNA Enzymatic Labeling kit and labeled with Cy-5. Cy3-labeled genomic DNA extracted from mouse spleen was used as a reference. Two micrograms of sample and reference DNA were hybridized to Agilent 44K whole-mouse genome comparative genomic hybridization (CGH) arrays (Agilent, G4414A). Samples were prepared using the direct method according to the manufacturer's protocol, to which minor changes were incorporated. Hybridization took place in a rotisserie oven for 72 h, set to 65°C and a rotation speed of 20 rpm [Scigene Rotator for 20 Agilent Surehyb chambers (part # 1070-20-0)]. The washing and scanning of the slides took place in an ozone-free area to prevent the degradation of the Cy5 dye. In turn, the slides were washed according to wash procedure B. After washing, the slides were dried and then scanned on an Agilent High-Resolution C scanner. Feature extraction was performed using Agilent software (FE 9.5.3.1).

Array-CGH data were processed using the R statistical framework with Bioconductor. The data were loaded and normalized as described in the snapCGH package (9), using the Edwards log linear interpolation method for background correction, a weighted median subtraction for normalizing within each array, and the processCGH function for final processing and ordering of the data. DNA copy number estimates were generated using circular binary segmentation. Genes were annotated and positioned using the Agilent Mouse Chip annotation package from AnnotationDbi

(Bioconductor). After segmentation, the two probes present in the Met transgene were removed according to the peak present in the A66 mammary fat pad (MFP) profile. To generate the whole-genome plot, we averaged the copy number estimates for each probe across all samples. To detect regions of copy number change, a *t*-test was performed comparing the MMTV-Met^{mt}; Trp53fl/+;Cre with the MFP samples. The null hypothesis (mean copy number is not significantly different) was rejected for probes with an false discovery rate (FDR)-adjusted *P* value less than 0.01. Segment plots were generated by plotting probes according to their genomic position and colored by their log 2 copy number change relative to reference. Ideograms for these plots were generated according to ideogram information downloaded from the UCSC Genomic Browser.

MicroRNA Microarrays. Snap-frozen tissue pieces were powdered and homogenized as described for gene expression profiling. Total RNA was then extracted using the miRNeasy Mini Kit (Qiagen). Total RNA was quality control–tested using the Agilent 2100 Bioanalyzer with the RNA 6000 Pico Kit and Small RNA Kit (both Agilent). Labeling and hybridization were carried out with the miRNA Complete Labeling and Hyb Kit and microRNA (miRNA) Spike-in kit (both Agilent) to single-channel arrays (Agilent 8 × 15K miRNA Oligo Microarray Kit, G4472A). Arrays were washed as directed by the manufacturer and scanned using an Agilent dual laser scanner (G2505B) Feature extraction was carried out using Agilent software (FE 10.7.3).

Mouse model miRNA array data were quantile normalized in the R statistical framework with Bioconductor.

To generate a human claudin-low miRNA signature, normalized gene and miRNA expression data from 207 paired breast tumors was obtained from the Buffa et al. publicly available dataset (Gene Expression Omnibus accession no. GSE22220) (10). The normalized intensity probes mapping to the same gene (National Center for Biotechnology Information Entrez gene identifier, as defined by the manufacturer) were averaged to generate independent expression estimates. Genes were median-centered and samples standardized to zero mean and unit variance. From the gene expression data, we identified claudin-low tumors by applying the previously published 9–cell line claudin-low predictor (11). Finally, a two-class unpaired significance analysis of microarrays was used to identify 53 miRNAs differentially expressed between claudin-low tumors versus others (false discovery rate < 4%).

Hierarchical clustering of the mouse samples was then performed using the miRNAs from the human claudin-low profile (described earlier) and Ward's agglomeration algorithm with an Euclidean distance metric. The statistical significance of the association of these miRNAs with the tumor clustering was made with gene set enrichment analysis. The GSEA background distribution was obtained using 10,000 random signatures of the same size.

Statistical Analysis of Clinical Outcomes in the Axillary Node-Negative Cohort. All of the analyses were conducted in a cohort of 618 [both Met protooncogene (MET) and tumor protein p53 (TP53) available] axillary node-negative human breast cancer cases (n = 42). Statistical analyses were performed using the SAS v9.2 statistical software program (SAS, Inc.). The Kaplan-Meier curve was produced using R statistical software version 2.15.0 (www.r-project.org). For all tests, alpha error was set at 5%.

Association Analysis of Combined MET and TP53 Tissue Microarray Markers with Clinical-Pathological Markers and the Tissue Microarray Markers Used to Define Subgroups. The χ^2 test or Fisher's exact test were used to analyze the associations. We compared frequency distribution of each marker in patients with tumors positive for both MET and TP53 with distribution in a combined group with tumors positive for neither or only one. Results are given in Dataset S1, Table S11.

Association Analysis of Combined MET and TP53 Tissue Microarray Markers with Subgroups. The basal group was characterized as human epidermal growth factor receptor 2 (Her2)− and estrogen receptor (ER)− and progesterone receptor (PR)− and either cytokeratin (CK)5+ or epidermal growth factor receptor (EGFR)+, and the triple-negative phenotype (TNP)-nonbasal was characterized by Her2− and ER− and PR− and CK5− and EGFR−. We investigated whether the expression of MET/TP53 correlates with molecular subtypes (basal and TNP-nonbasal subtypes) using a χ^2 test. Results are given in Table 2.

DFS Analysis of Combined MET and TP53 Protein Levels. Analyses of the association of disease-free survival (DFS) with MET protein status were conducted using Kaplan-Meier plots and a standard Cox proportional hazards model with and without including traditional clinicopathological factors as covariates (multivariate and univariate models, respectively) (Dataset S1, Table S10). The traditional factors used were menopausal status, tumor size, histological grade, estrogen receptor status, lymphatic invasion, age at diagnosis, and adjuvant treatment received. To assess the association of DFS with the MET and TP53 protein status jointly, we compared survival of patients with tumors positive for both MET and TP53 with that of a combined group with tumors positive for neither or only one of MET and TP53 (Kaplan-Meier Fig. 7 for MET and TP53; Dataset S1, Table S12), adjusting for the same traditional factors.

Patients with tumors expressing high levels of both MET and TP53 (MET+/TP53+; n1=86, n2=19) show reduced DFS in comparison with the other groups (MET+/TP53-; n1=189, n2=22, MET-/TP53+;n1=56, n2=7, and MET-/TP53-; n1=287, n2=28, where n1 is the number of cases and n2 is the number of recurrences) (KM Fig. 7 for MET and TP53, log-rank P=1.20e-03).

The association of MET status and DFS became nonsignificant [relative risk (RR), 1.35; 95% confidence interval (CI), 0.87–2.10; P = 0.1851] at the 5% significance level in the multivariate model, although it was significant in the univariate model (RR, 1.57; 95% CI, 1.02–2.42; P = 0.0411) (Dataset S1, Table S12). Remarkably, when MET and TP53 were considered jointly, we found a 2-fold elevated risk of disease recurrence when the tumor specimen had both MET and TP53 compared with those having only one or neither of the proteins (RR, 2.04; 95% CI, 1.15–3.62; P = 0.0149) (KM Fig. 7 for MET and TP53; Dataset S1, Table S12).

Real-Time PCR. All primers were designed using Primer3 software (available at frodo.wi.mit.edu); sequences are shown in Dataset S1, Table S14.

Reverse transcription was performed using the Roche reverse transcription kit (Roche Transcriptor First-Strand cDNA Synthesis Kit). Real-time PCR was carried out using LightCycler 480 SYBR Green I Master reagents (Roche) and a Roche LightCycler 480.

Data were normalized to 3 housekeeping genes (*Gapdh*, *Rpl13a*, and *Hprt*), using a normalization factor generated in geNorm software (BioGazelle) (12). Averaged PCR data for biological replicates (five MMTV-Met;Trp53fl/+;Cre, five Trp53fl/+;Cre, three MMTV-Met solid, and three MMTV-Met mixed-pathology tumors) are presented.

Genotyping PCR for *Trp53*. Genomic DNA was extracted from tumor cells isolated from MMTV-Met;Trp53fl/+;Cre and MMTV-Met mammary tumors, as described earlier. This avoided contaminating signal from tumor stromal components. The strategy for PCR detection of wild-type and recombined *Trp53* alleles was based on that published by Jonkers and colleagues, who generated the *Trp53*-floxed mice (13). PCR primer designs were as follows:

p53|1F 5′ggttaaacccagcttgacca 3′; p53|1R 5′cgaggcttgtcccaactcta 3′ and

p53|10F 5′aaaaccccaccctgctagat 3′; p53|10R 5′tgggtagggatattcacagaaca 3′.

The following PCR cycling conditions were used: 95°C 1 min, 95°C 10 s, 58°C 5 s, 72°C 1 s, 72°C 30 s (×33 cycles).

Western Blotting. Snap-frozen mammary tumors and normal mammary gland samples were powdered under liquid nitrogen and then lysed for protein extraction using a 1% Triton lysis buffer (50 mM Hepes at pH 7.5, 150 mM NaCl, 1.5 mM $MgCl_2$, 1 mM EGTA, 10% glycerol, 1% Triton X-100, 1 mM phenylmethylsulfonyl fluoride, 1 mM sodium vanadate, 1 mM sodium fluoride, 10 μg/mL aprotinin, and 10 μg/mL leupeptin).

Protein lysates from tumor-derived cell lines were generated using TNE lysis buffer (50 mM Tris at pH 8.0, 150 mM NaCl, 1% Nonidet P-40, 2 mM EDTA at pH 8.0, 1 mM phenylmethylsulfonyl fluoride, 1 mM sodium vanadate, 1 mM sodium fluoride, 10 μg/mL aprotinin, and 10 μg/mL leupeptin).

Proteins were resolved by SDS/PAGE and transferred to a nitrocellulose or polyvinylidene difluoride (PVDF) membrane. Membranes were blocked in 2% milk (Cldn1 detection) or Odyssey Blocking Buffer (LI-COR Biosciences) (Met and pMet detection) for 1 h at room temperature and probed with primary antibody (diluted in 2% milk for Cldn 1 or Odyssey Blocking Buffer for Met and pMet) overnight at 4 °C. Membranes were washed 3 times in tris buffered saline with tween 20 (TBST) and incubated with HRP-conjugated (Cldn1) or fluorophore-conjugated (Met, pMet) secondary antibodies for 1 h at room temperature. After washing 3 times in TBST, bound proteins were detected with an ECL-kit (Amersham Biosciences) or by scanning with the LI-COR Odyssey (LI-COR Biosciences), as appropriate. Primary and secondary antibodies are detailed in Dataset S1, Table S13. Quantification shown in Fig. S4 was performed relative to actin, using Odyssey V3 software.

Tail Vein Injection with Luciferase-Expressing Primary Cells. Primary cells isolated from an MMTV-Met;Trp53fl/+;Cre spindloid mammary tumor were transduced with the pLenti PGK V5-LUC Neo lentivirus (Addgene plasmid 21471) encoding firefly luciferase and originally made by Eric Campeau (University of Massachusetts, Worcester, MA) (14). Cells with stable expression of the gene were selected under G418.

Athymic nude mice (Taconic Farms, Inc.) (n = 35) were injected via the tail vein with 0.5×10^6 cells. Luciferase activity in the lungs was confirmed by imaging immediately postinjection. For imaging, mice were injected (intraperitoneally) with the luciferase substrate D-luciferin (Caliper Life Sciences, Inc.) dissolved in PBS (50 μL at 30 mg/mL), anaesthetized with isoflurane, and imaged by bioluminescence (Xenogen IVIS 100, Caliper Life Sciences, Inc.) at 1-min intervals over 10 min. Mice were imaged twice per week thereafter to monitor the development of metastases.

Treatment of Mice with Crizotinib. Mice were administered Crizotinib (LC Laboratories) by oral gavage (45 $mg \cdot kg^{-1} \cdot d^{-1}$, dissolved in water) daily from the day of tail vein injection. A control group of 15 mice was gavaged with water only.

Metastasis Scoring. Metastatic lesions in the lungs and livers of tail vein–injected mice were scored by counting the number of lesions across four step sections each of 50 μm. Lesions that were present in multiple steps were counted only once.

In vitro Proliferation Assays. Cell lines were seeded in 12-well plates (one plate per time) at 60,000 cells per well. At each 24-h point, cells were trypsinized and counted using an automated cell counter (Cellometer, Nexcelom Bioscience). A total of four cell lines were used, and the assay was performed in duplicate using both PHA665752 (1 μM) and Crizotinib (1 μM).

Soft Agar Assays. Soft agar assays were performed over a period of 10 d, as previously described (15), seeding 30,000 cells per well in six-well plates. Colonies were imaged for scoring by size using Infinity Analyze Software (Lumenera Corp.). Whole-well images were taken using the Zeiss Axio Zoom V16 microscope. A total of four cell lines were assayed, and the assay was performed in duplicate using both PHA665752 (1 μM) and Crizotinib (1 μM).

Flow Cytometry. Four cell lines were treated with PHA665752 (1 μM) or Crizotinib (1 μM) for 48 h before labeling and flow cytometry, compared with untreated cells and cells treated with DMSO for the same period. For flow cytometry, the medium containing floating cells was collected and combined with adherent cells that were trypsinized from the plates. A total of 1 million cells were then labeled using the Annexin-V FLUOS staining kit (Roche), according to the manufacturer's instructions. Flow cytometry was performed on the LSRII (BD Biosciences).

1. Ponzo MG, et al. (2009) Met induces mammary tumors with diverse histologies and is associated with poor outcome and human basal breast cancer. *Proc Natl Acad Sci USA* 106(31):12903–12908.
2. Durinck S, Spellman PT, Birney E, Huber W (2009) Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols* 4:1184–1191.
3. Herschkowitz JI, et al. (2007) Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol* 8(5):R76.
4. Smyth GK (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 3:Article3.
5. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc Ser A Stat Soc* 57(1):289–300.
6. Subramanian A, et al. (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 102(43):15545–15550.
7. Mootha VK, et al. (2003) PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* 34(3): 267–273.
8. Culhane AC, et al. (2010) GeneSigDB - a curated database of gene expression signatures. *Nucleic Acids* 38:D716–726.
9. Smith, ML, Marioni, JC, Hardcastle, TJ, Thorne, NP (2006) snapCGH: Segmentation, normalization and processing of aCGH data users' guide. *Bioconductor*.
10. Buffa FM, et al. (2011) microRNA-associated progression pathways and potential therapeutic targets identified by integrated mRNA and microRNA expression profiling in breast cancer. *Cancer Res* 71(17):5635–5645.
11. Prat A, et al. (2010) Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res* 12(5):R68.
12. Vandesompele J, et al. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 3(7):RESEARCH0034.
13. Jonkers J, et al. (2001) Synergistic tumor suppressor activity of BRCA2 and p53 in a conditional mouse model for breast cancer. *Nat Genet* 29(4):418–425.
14. Campeau E, et al. (2009) A versatile viral system for expression and depletion of proteins in mammalian cells. *PLoS ONE* 4(8):e6529.
15. Mak HH, et al. (2007) Oncogenic activation of the Met receptor tyrosine kinase fusion protein, Tpr-Met, involves exclusion from the endocytic degradative pathway. *Oncogene* 26(51):7213–7221.

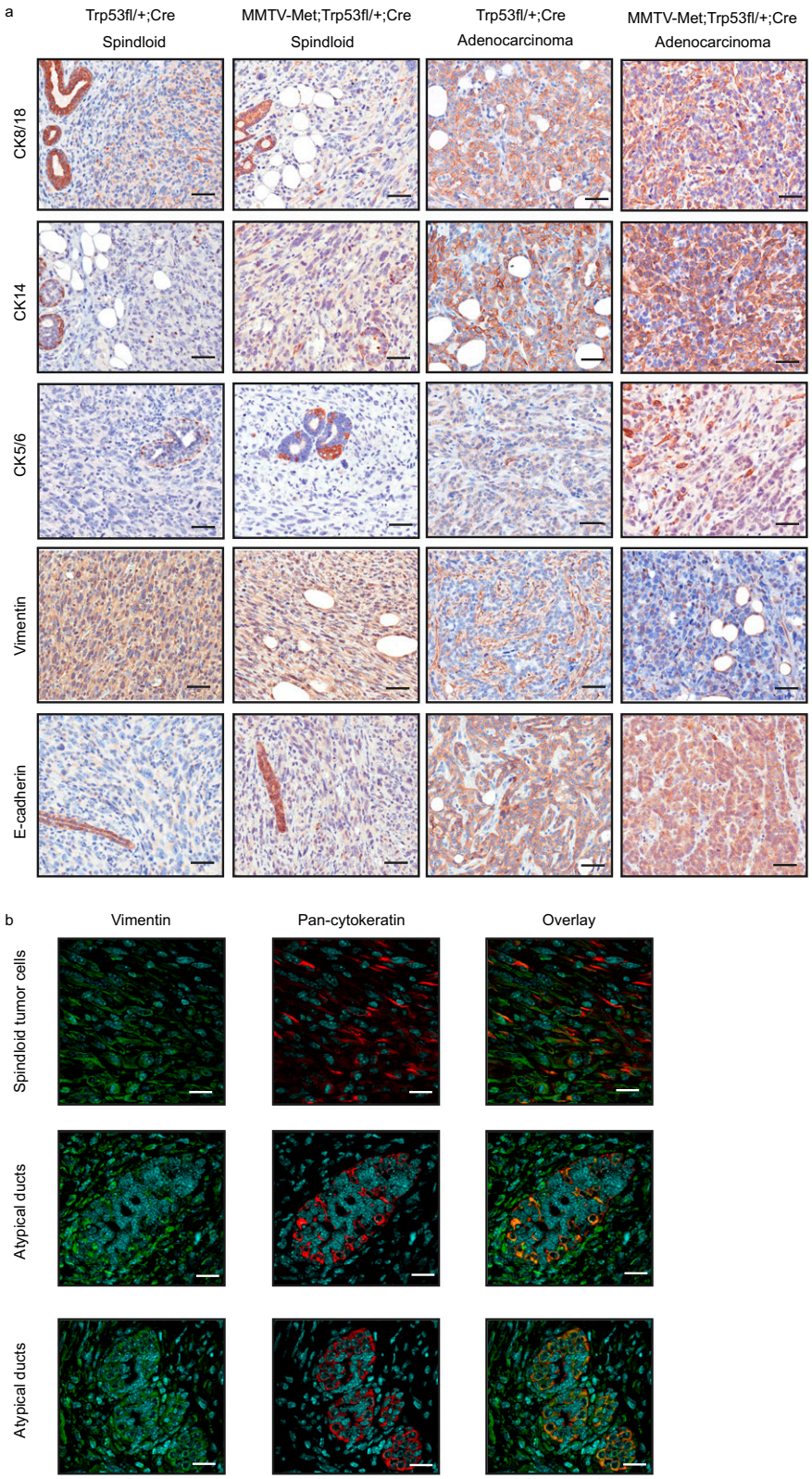


Fig. S1. Immunohistochemical staining patterns of MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre spindloid tumors are consistent with an epithelial-to-mesenchymal transition (EMT). A panel of MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre tumors were stained with antibodies for cytokeratins (CKs) and E-cadherin, typically expressed by epithelial cells (*A*). Expression of these markers in tumors of spindloid pathology was sporadic, and in the majority of tumors it was localized to ductal structures. In contrast, tumors of adenocarcinoma pathology stained strongly for CK14 and 8/18 and also contained pockets of cells positive for CK5/6. These tumors were also positive for E-cadherin. Spindloid tumor cells stained positively for the mesenchymal marker vimentin, whereas in adenocarcinomas this was localized only to tumor-infiltrating stromal cells. (Scale bars, 50 μm.) Spindle tumor cells in MMTV-Met^{mt};Trp53fl/+;Cre tumors also showed colabeling with antibodies directed against pan-cytokeratin (red) and vimentin (green), supportive of an EMT (*B*). Cytokeratin-positive ductal cells also label positive for vimentin, thus capturing the early phases of EMT within epithelium. (Scale bars, 20 μm.)

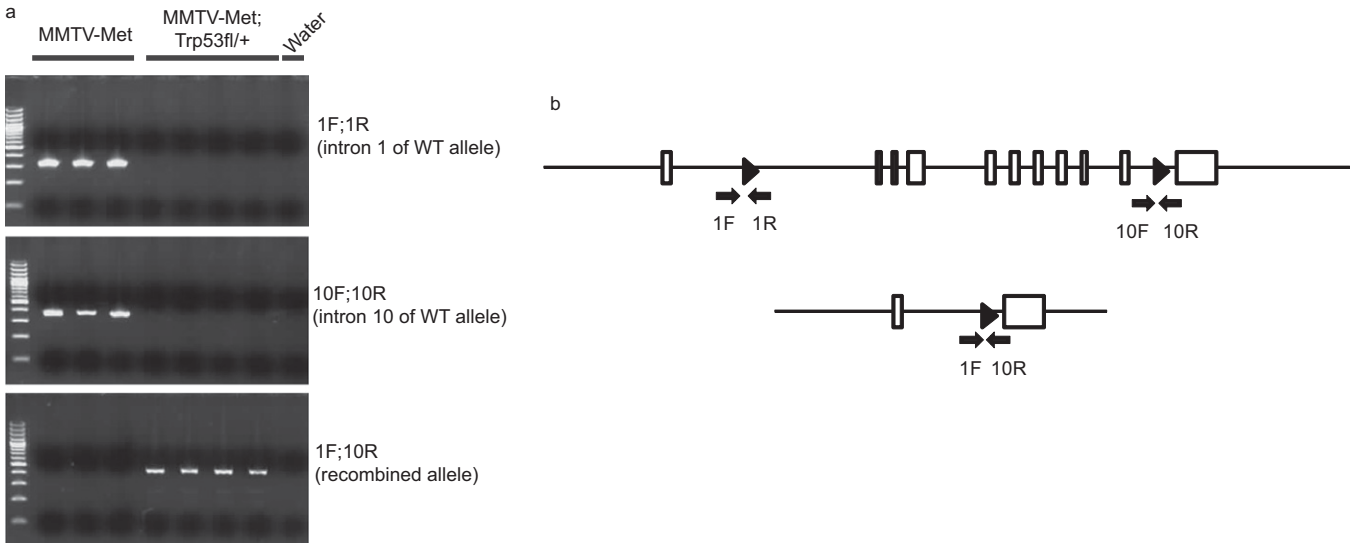


Fig. S2. MMTV-Met^{mt};Trp53fl/+;Cre tumors undergo loss of heterozygosity (LOH) at the Trp53fl/+ locus. DNA from MMTV-Met^{mt} and MMTV-Met^{mt};Trp53fl/+;Cre primary tumor cells was used in PCR with primers that detected both wild-type and Cre-recombined *Trp53* alleles (*A*). In MMTV-Met^{mt};Trp53fl/+;Cre mice, one *Trp53* allele contains locus of X-over P1 (LoxP1) sites (►) in introns 1 and 10, such that Cre-mediated recombination results in excision of exons 2–10 (*B*). Primers located in introns 1 and 10 (1F:1R or 10F:10R) will only generate PCR product if an unrecombined *Trp53* allele is present, as shown for MMTV-Met^{mt} tumor cells (*A*). Absence of these PCR products in MMTV-Met^{mt};Trp53$^{+/-}$;Cre tumor cells indicates that the wild-type (unfloxed) allele is also missing, demonstrating LOH. PCR using primers 1F and 10R generates the small product that results from the Cre-mediated recombination of the floxed allele (*A*). Adapted from figure 2 of Jonkers et al., *Nature Genetics* 2001.

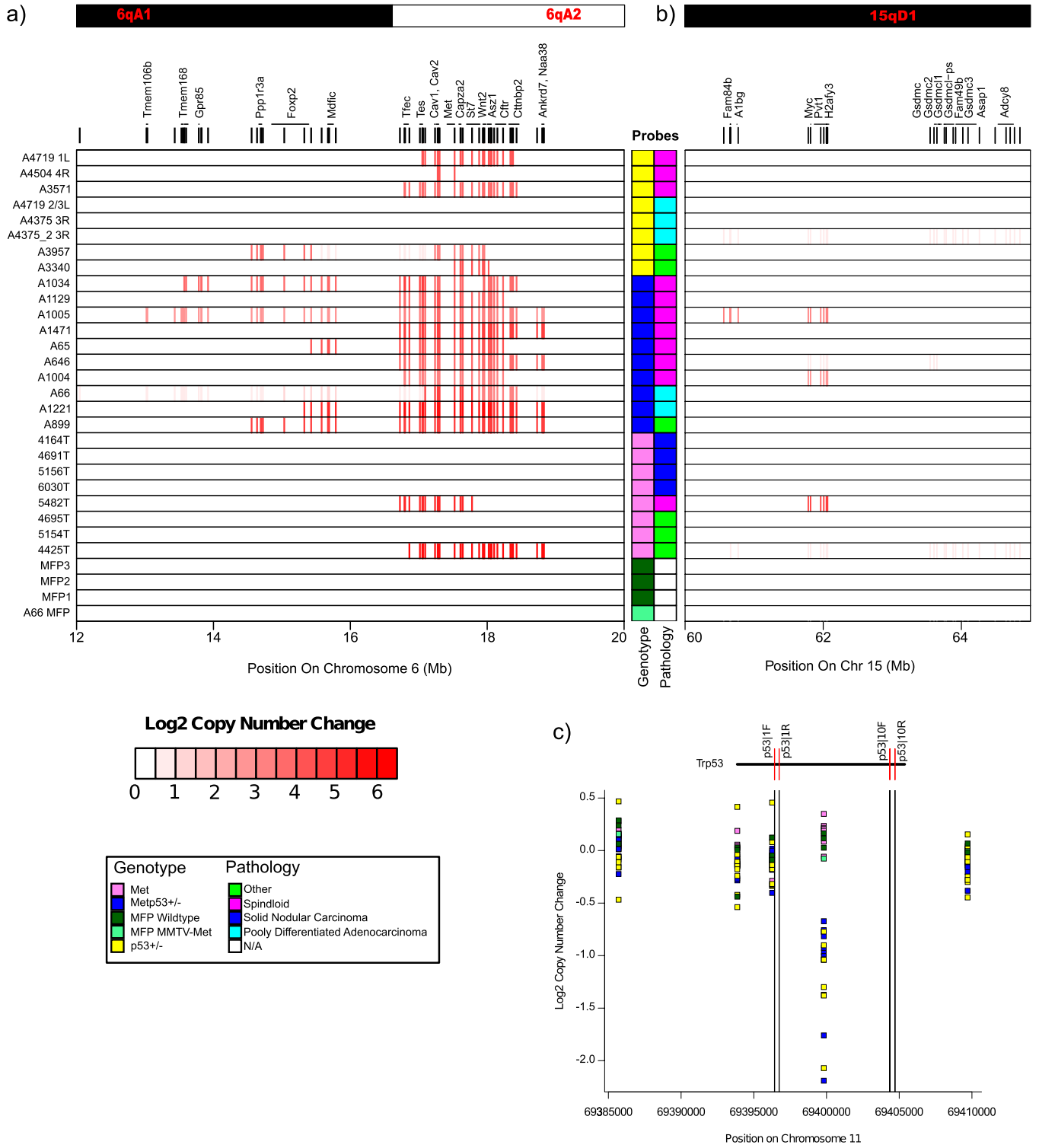


Fig. S3. Genomic amplification of Met is detected in all MMTV-Met^{mt};Trp53fl/+;Cre tumors and in Trp53fl/+;Cre tumors of spindloid pathology. Array-CGH on 10 MMTV-Met^{mt};Trp53fl/+;Cre, eight Trp53fl/+;Cre, and eight MMTV-Met^{mt} tumors showed that genomic amplification of Met and immediately adjacent loci such as Cav1 occurred in 10 of 10 MMTV-Met^{mt};Trp53fl/+;Cre tumors, five of eight Trp53fl/+;Cre tumors (all those with spindloid pathology), and two of eight MMTV-Met^{mt} tumors (one of which was spindloid) (*A*). Other genomic events included amplification of Myc in three of 10 MMTV-Met^{mt};Trp53fl/+;Cre tumors, one of eight Trp53fl/+;Cre tumors, and two of eight MMTV-Met^{mt} tumors (*B*). Array CGH also confirmed LOH at the Trp53 locus in all MMTV-Met^{mt};Trp53fl/+;Cre and all Trp53fl/+;Cre tumors (*C*).

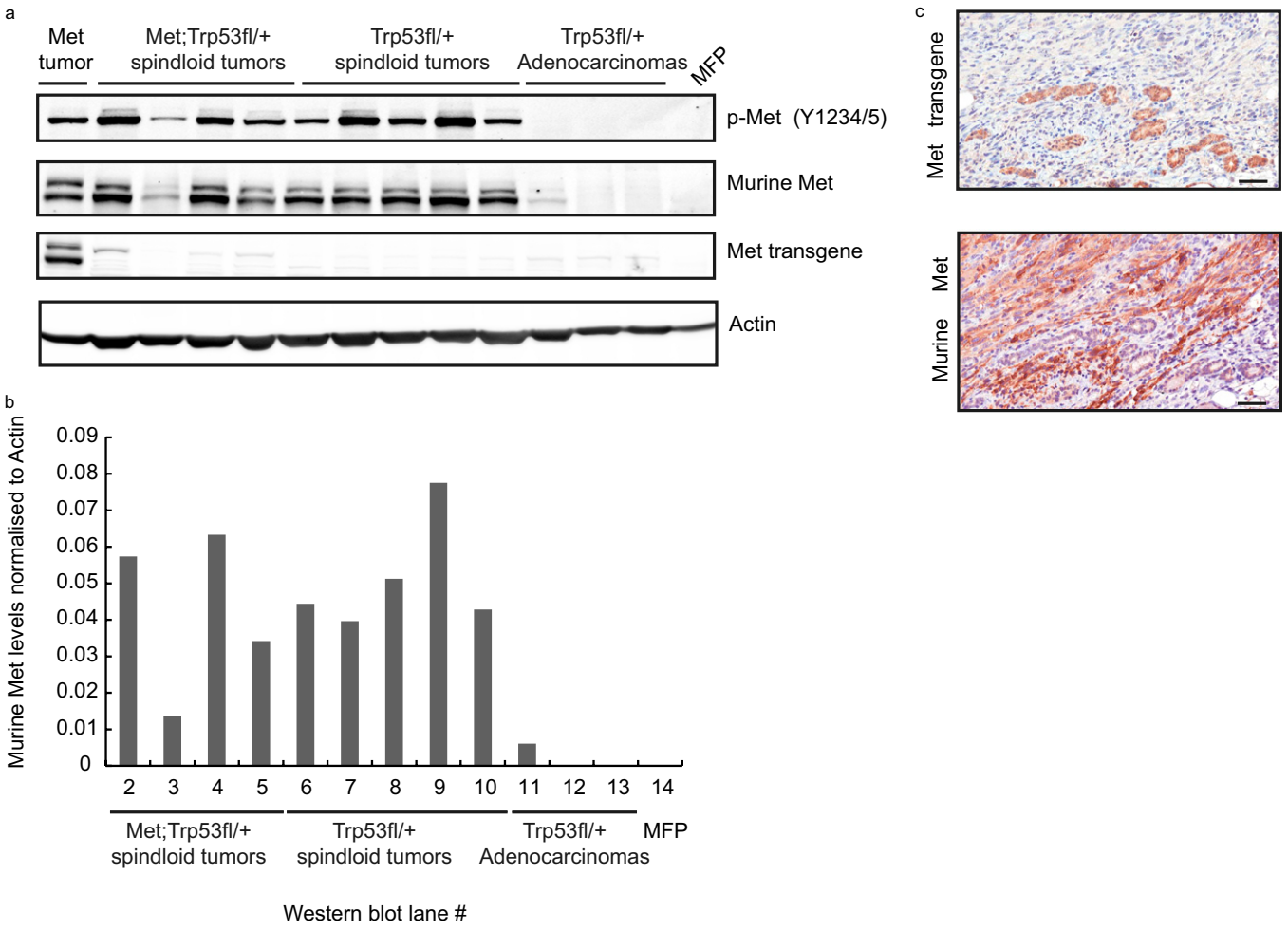


Fig. S4. MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre spindloid tumors express elevated levels of endogenous murine *Met*. Immunoblotting confirmed that genomic amplification of *Met* results in an increase in MET protein levels in MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre tumors of spindloid pathology (*A*). Use of a p-MET (Y1234/1235) antibody confirms that the murine MET protein is highly activated (*A*). Similar levels of MET activation are also seen in Trp53fl/+;Cre spindloid tumors (lanes 6–10), but not Trp53fl/+;Cre adenocarcinomas (lanes 11–13), supporting a role for MET in promoting a spindloid pathology. Protein from a normal MFP (lane 14) is included as a control. Quantification of the immunoblot for murine MET (relative to the Actin loading control) was performed using Odyssey V3 software (LI-COR Biosciences) (*B*). In addition, although the MMTV-Met^{mt} transgene protein was detected in a control MMTV-Met^{mt} solid carcinoma with wild-type *Trp53* (lane 1), MMTV-Met^{mt};Trp53fl/+;Cre spindloid tumors (lanes 2–5) showed repression of the MET transgene (*A*). Transgene switch-off and expression of endogenous murine MET was also confirmed by immunohistochemistry, with which transgenic MET could be detected in normal mammary glands but not in tumor cells that had undergone EMT, which instead expressed high levels of murine MET protein (*C*).

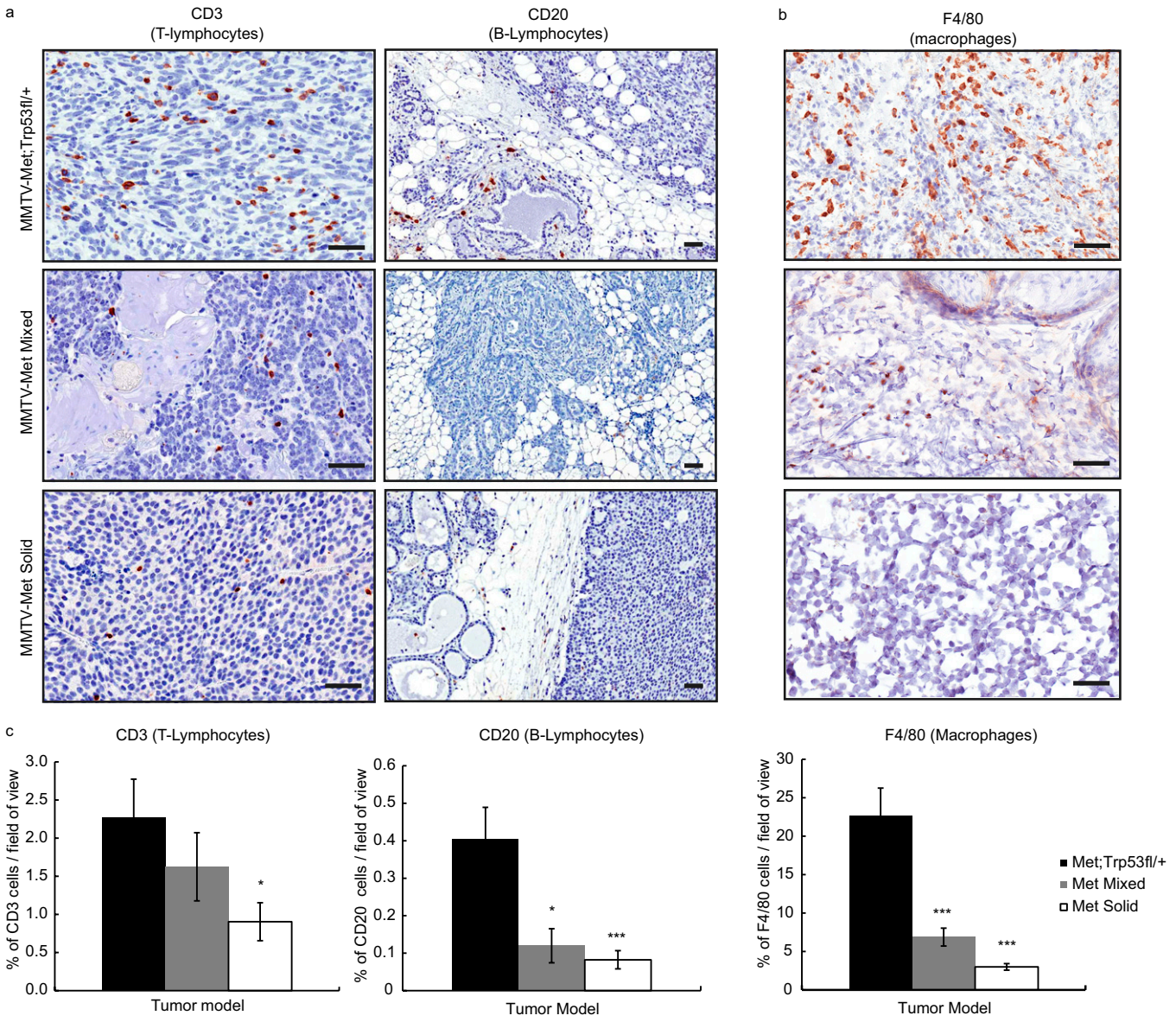


Fig. S5. MMTV-Met^{mt};Trp53fl/+;Cre tumors contain a high degree of lymphocytic and macrophage infiltration relative to MMTV-Met^{mt} tumors. The degree of T- and B-lymphocyte infiltration in MMTV-Met^{mt};Trp53fl/+;Cre and MMTV-Met^{mt} tumors was investigated by immunohistochemistry using CD3 and CD20 antibodies, respectively (*A*). Macrophage infiltration was assessed by immunostaining for F4/80 (*B*). In each case, the number of positive cells was counted using an algorithm in the program ImageScope (Aperio Technologies) and expressed as a percentage of all cells per field of view; 14 fields of view were counted, and a minimum of 3 tumors per tumor type were used (*C*). MMTV-Met^{mt};Trp53fl/+;Cre tumors contained significantly more infiltrating T lymphocytes than MMTV-Met^{mt} solid tumors ($P = 0.044$), and T lymphocytes were largely restricted to the adjacent stroma. In all tumors, B-lymphocytes were only detected at tumor peripheries, but they were detected at significantly higher numbers in MMTV-Met^{mt};Trp53fl/+;Cre tumors than in MMTV-Met^{mt} mixed- and solid-pathology tumors ($P = 0.015$ and 0.007, respectively). Macrophage infiltration was significantly higher in MMTV-Met^{mt};Trp53fl/+;Cre tumors compared with MMTV-Met^{mt} mixed- and solid-pathology tumors ($P = 0.002$ and 0.003, respectively). (Scale bars, 50 μm.)

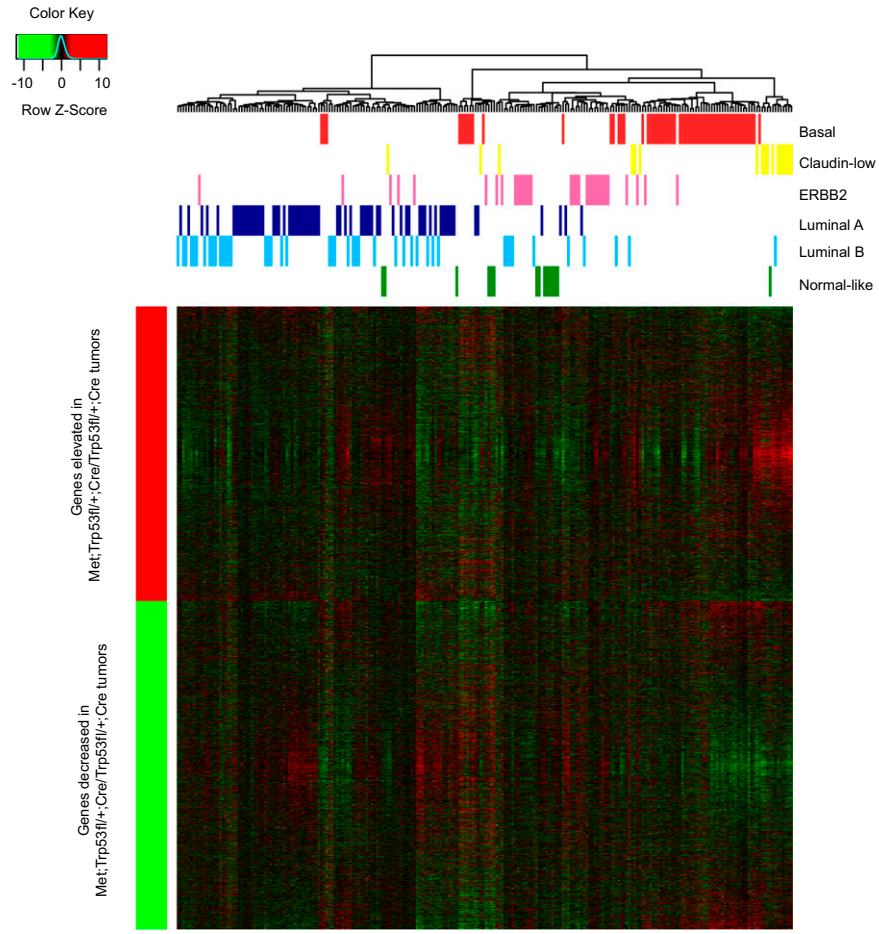


Fig. S6. Identification of the human claudin-low molecular subtype through application of the mouse gene expression signature. Genes differentially expressed between MMTV-Met^{mt};Trp53fl/+;Cre or Trp53fl/+;Cre spindloid tumors and MMTV-Met^{mt} tumors were obtained and orthologs applied to a human breast cancer dataset. Hierarchical clustering revealed that the claudin-low subtype of breast tumors group together with a distinct molecular profile that resembles murine MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+ spindloid tumors. Gene-set enrichment analysis revealed that this association was highly significant ($P < 0.0001$).

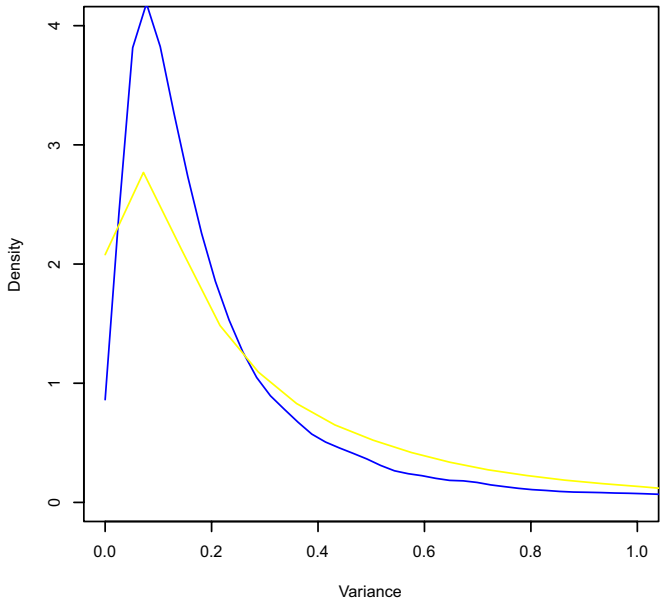


Fig. S7. MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre tumors of spindloid pathology show varying degrees of heterogeneity. Lines represent the distribution of gene variances over all genes on the microarray. The distribution for Trp53fl/+;Cre spindloid tumors (yellow) is significantly greater than that for MMTV-Met^{mt};Trp53fl/+;Cre tumors (blue) ($P < 2.2 \times 10^{-6}$).

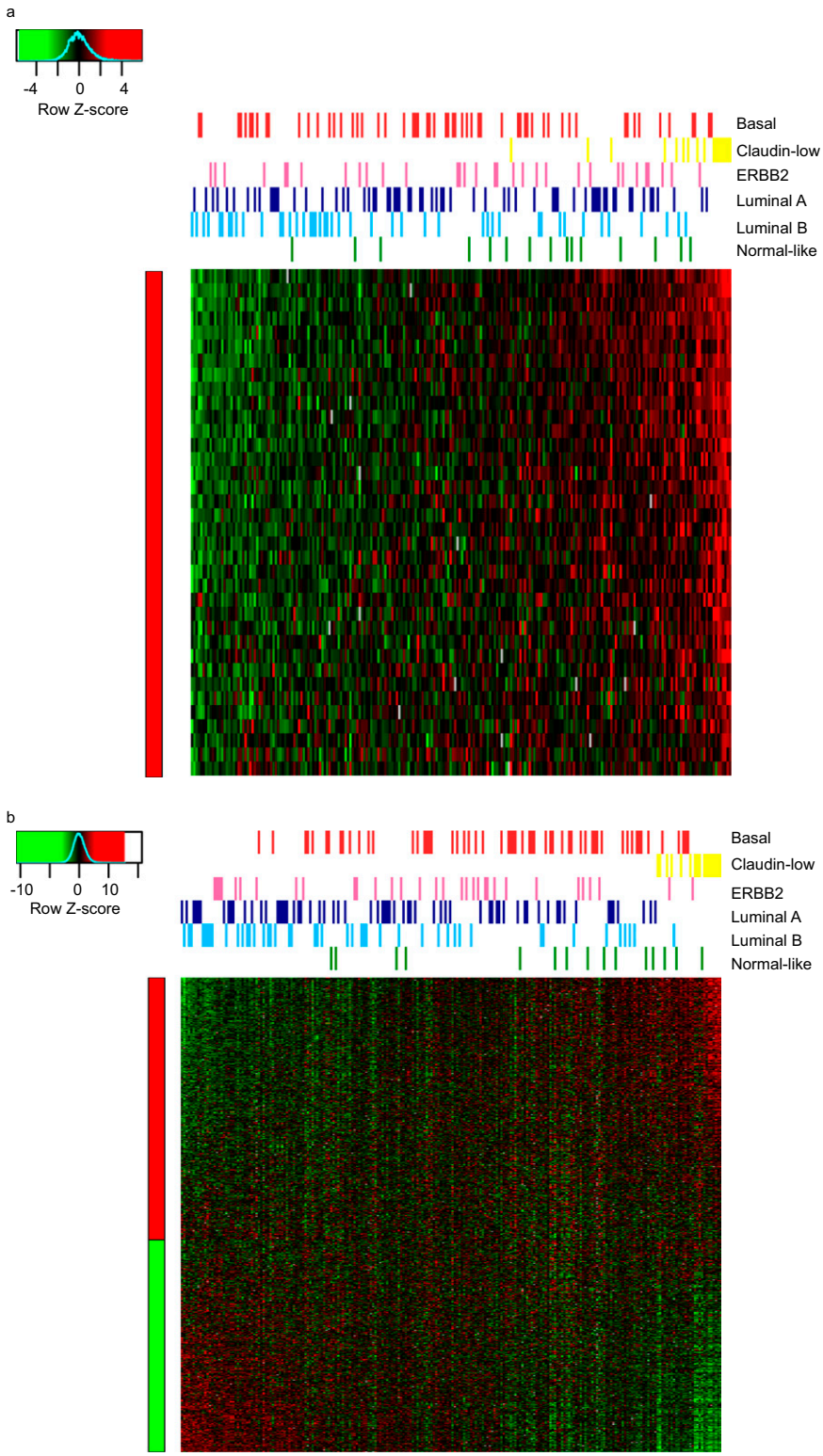


Fig. S8. The 36-gene intersect identifies claudin-low patients with an equivalent degree of accuracy as the published signature of 777 genes. Heat map of human breast tumors using the 36-gene intersect (*A*) and a previously published claudin-low signature (*B*). Tumors were linearly ordered from left to right, representing less to greater expression of each signature, respectively. Tumors classified as claudin-low consistently order to the right of the heat maps, signifying that both signatures are exclusively associated with this subtype. This association was highly significant by GSEA ($P < 0.0001$ for both signatures).

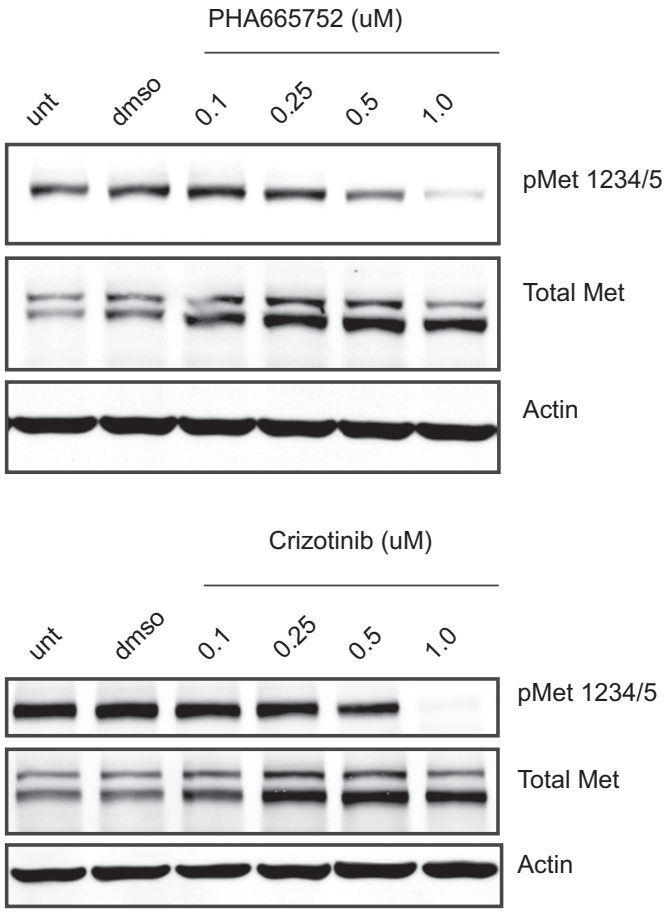


Fig. S9. Titration of MET kinase inhibitors PHA665752 and Crizotinib on spindloid MMTV-Met^{mt};Trp53fl/+;Cre tumor cells with *Met* amplification. A range of concentrations of MET inhibitor (PHA665752, *Upper*; Crizotinib, *Lower*) were tested on MMTV-Met^{mt};Trp53fl/+;Cre tumor cell lines to ensure effective inhibition of MET in assays presented in Figs. 4 and 5.

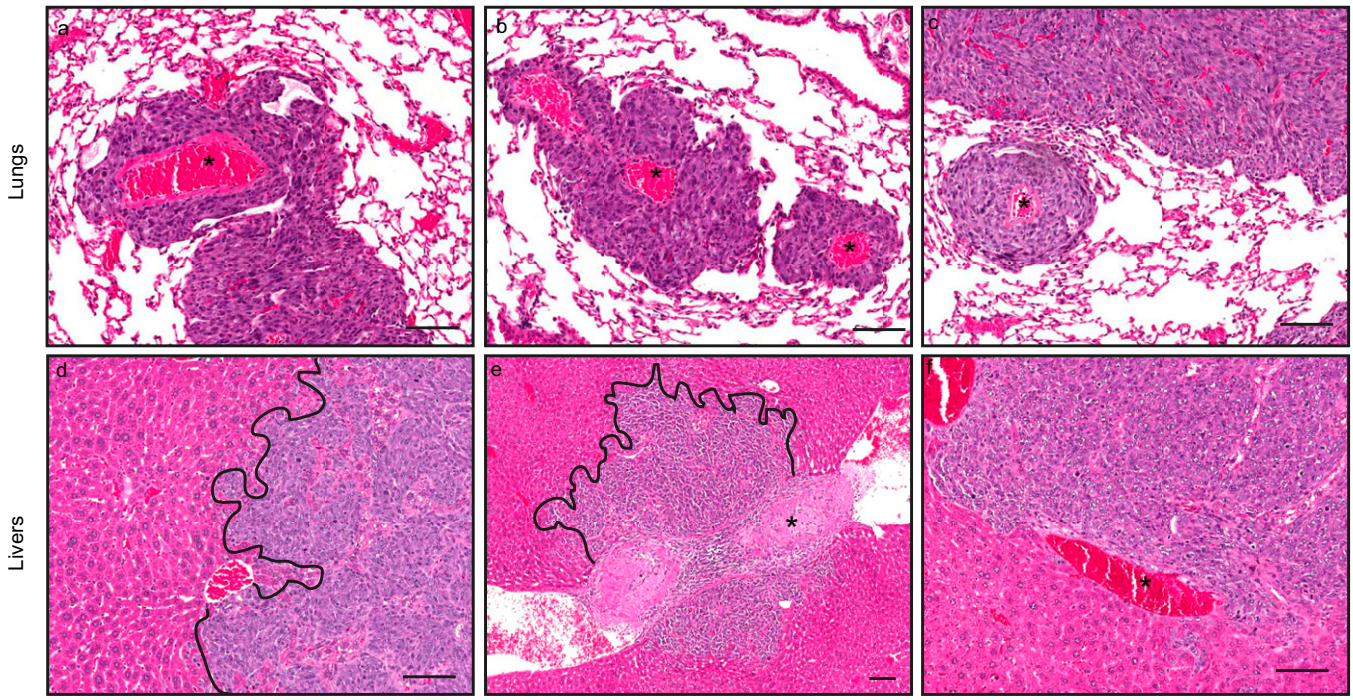


Fig. S10. Examples of lung and liver metastases in 3 nude mice injected i.v. with luciferase-expressing MMTV-Met;Trp53fl/+;Cre spindloid tumor cells. Twenty-four days after tail vein injection, mice showed extensive metastatic burden, as visualized by luminescence imaging (Fig. 6). Histological examination of lung (*A*–*C*) and liver (*D*–*F*) metastatic lesions showed growth emanating from blood vessels (*) and within the tissue bulk (rather than intravascular growth), which is evidence of extravasation. The invasive property of these cells is also illustrated by the pushing borders at the perimeter of lesions (examples outlined in *d* and *e*).

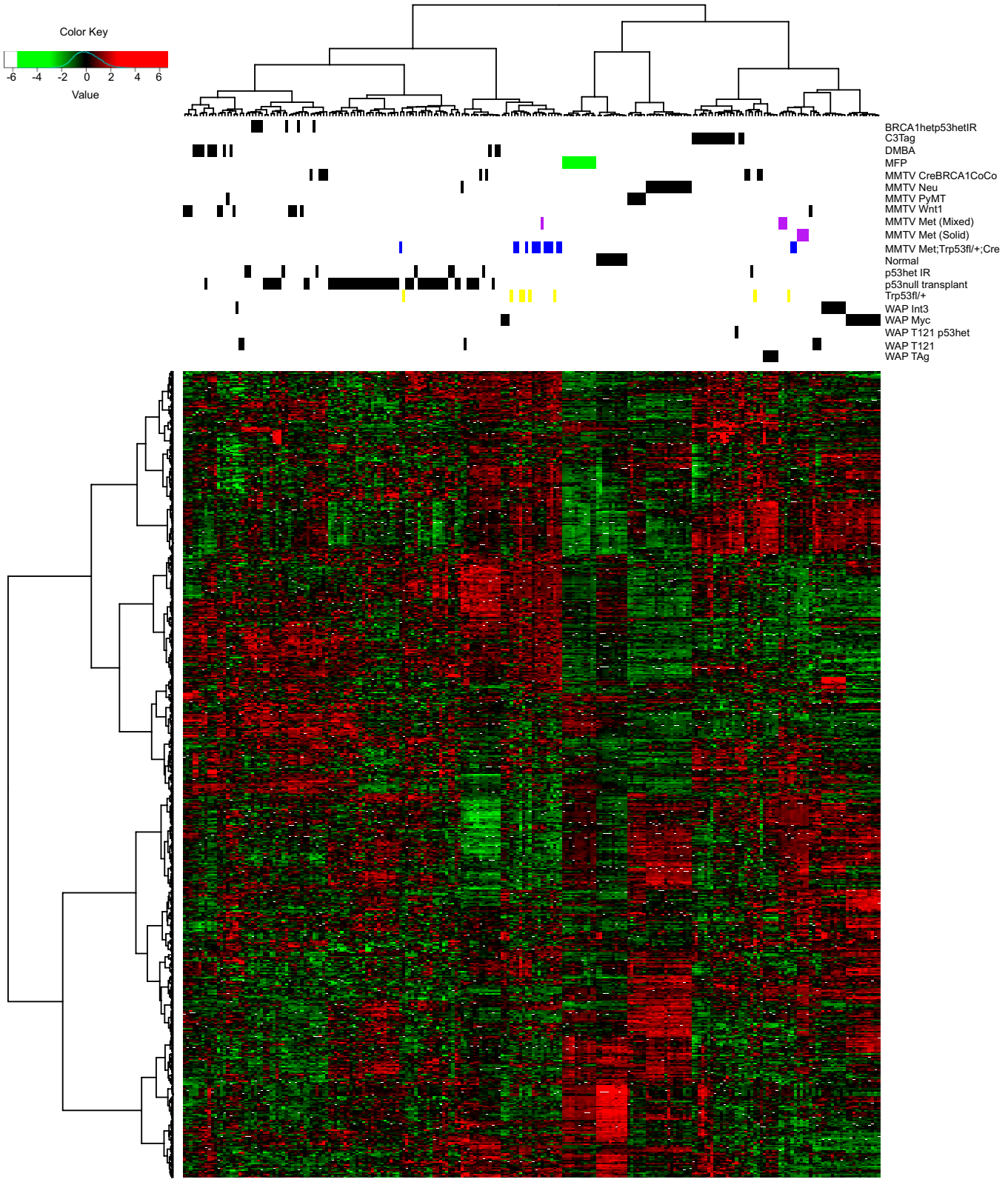


Fig. S11. MMTV-Met^{mt};Trp53fl/+;Cre spindloid tumors cluster with other mouse models that display an EMT phenotype. Unsupervised hierarchical clustering of gene expression data showed that MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre tumors of spindloid pathology group together and most closely to other mouse models in which a subset of tumors are documented to express an EMT-phenotype, such as DMBA, MMTV-Cre;Brca1$^{co/co}$, p53 null transplant, and WAP-Myc. Notably, although the majority (80%) of MMTV-Met^{mt};Trp53fl/+;Cre tumors display EMT pathology, it is clear from the heat map that only a small fraction of tumors from other models also show this phenotype.

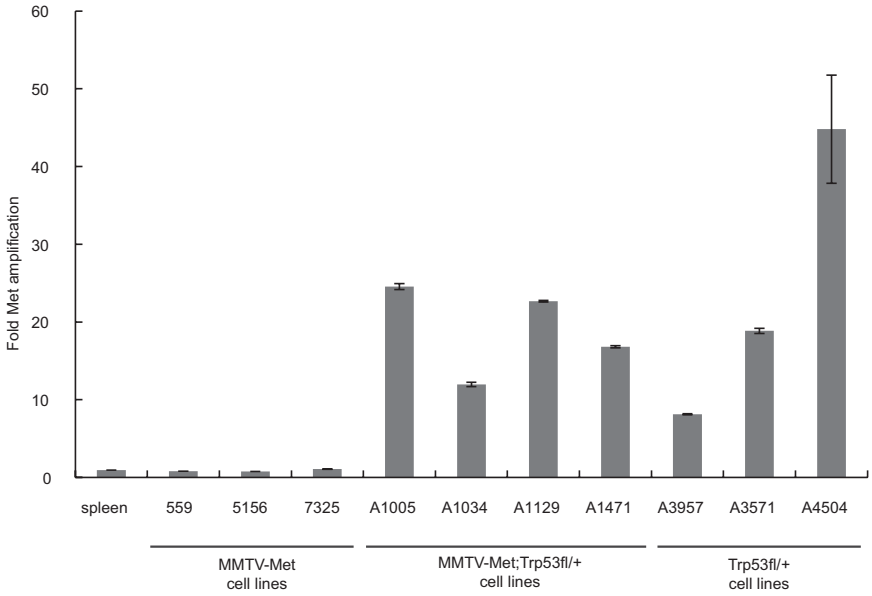


Fig. S12. The *Met* amplicon is retained in cell lines derived from MMTV-Met^{mt};Trp53fl/+;Cre and Trp53fl/+;Cre spindloid mammary tumors. Quantitative real-time PCR for *Met* gene copy number was performed on genomic DNA isolated from MMTV-Met^{mt}, MMTV-Met^{mt};Trp53fl/+;Cre, and Trp53fl/+;Cre tumor-derived cell lines, which had been cultured up to passage 20. Although MMTV-Met^{mt} tumor cell lines have an equivalent *Met* copy number to a wild-type spleen control, both MMTV-Met^{mt};Trp53fl/+;Cre (n = 4) and Trp53fl/+;Cre (n = 3) spindloid cell lines show elevated levels of genomic DNA encoding *Met*, demonstrating retention of the amplicon in culture. PCRs were performed in triplicate. Error bars, SEM.

Other Supporting Information Files

Dataset S1 (XLSX)

Supporting Data

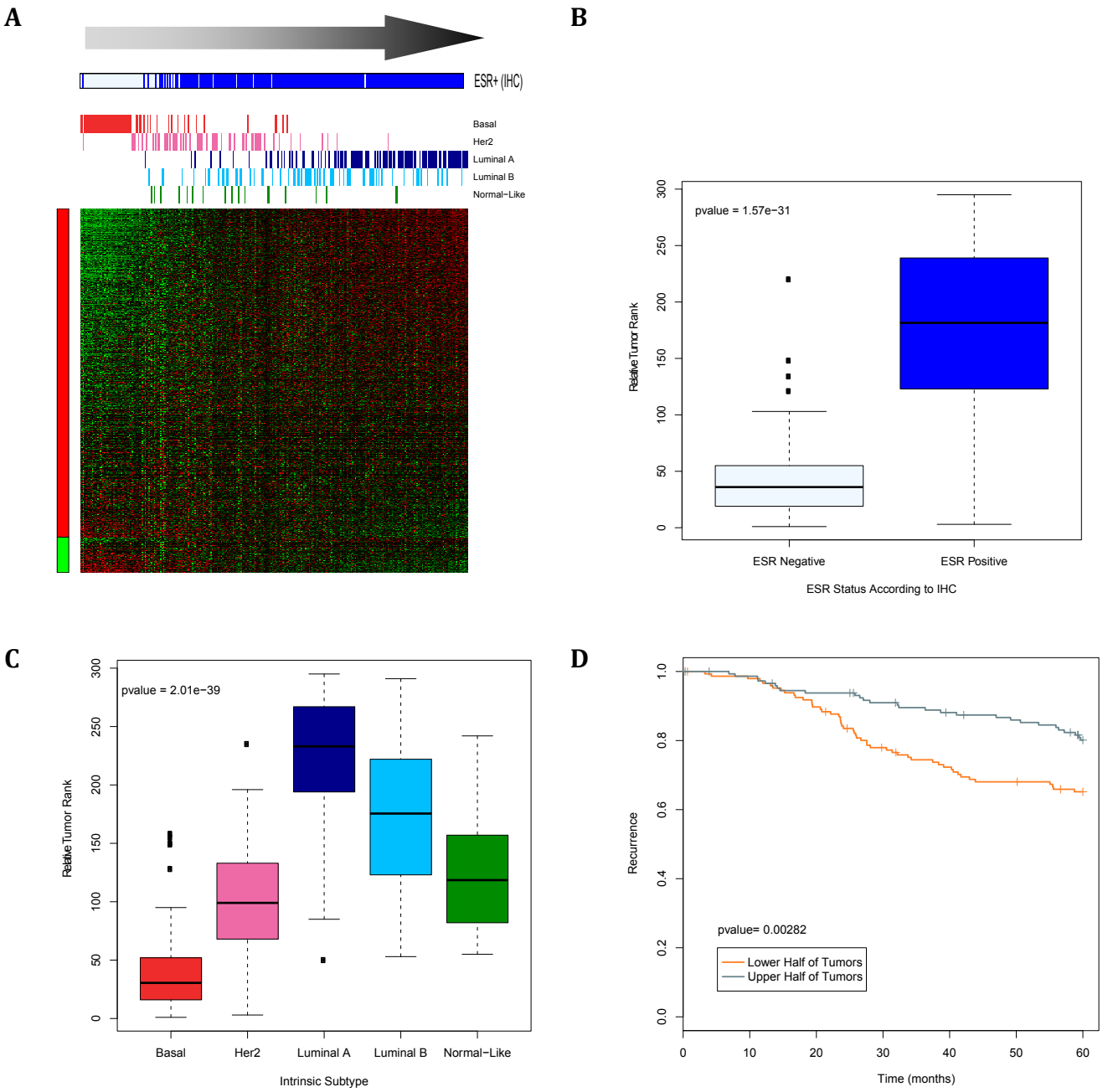


Figure 1. ESR activation signature in a breast cancer dataset. (A) Heatmap of ESR activation signature, with rows representing genes, and columns representing tumors. Gene expression is colored from green (low) to red (high). Samples are ordered from left (least ESR signaling activation) to right (most ESR signaling activation) using BreSAT. Arrow indicates increasing signature activation in the tumors. Patients are labeled according to their ESR IHC status (blue=positive), and their intrinsic subtype. (B) Patients ranks of the ESR- and ESR+ classes are displayed as boxplots, and are significantly different (p-value=1.6×10^{-31}). (C) Patient ranks of the intrinsic subtypes are displayed as boxplots, and are significantly different (p-value=2.0×10^{-39}). (D) Tumors were broadly divided in half according to their ranks, and Kaplan-Meier curve shows tumor recurrence of the two groups. The tumors with less ESR signaling activation have significantly worse outcome (p-value=2.8×10^{-3}). Expression data was obtained from [20]; ESR activation signature was obtained from [21].

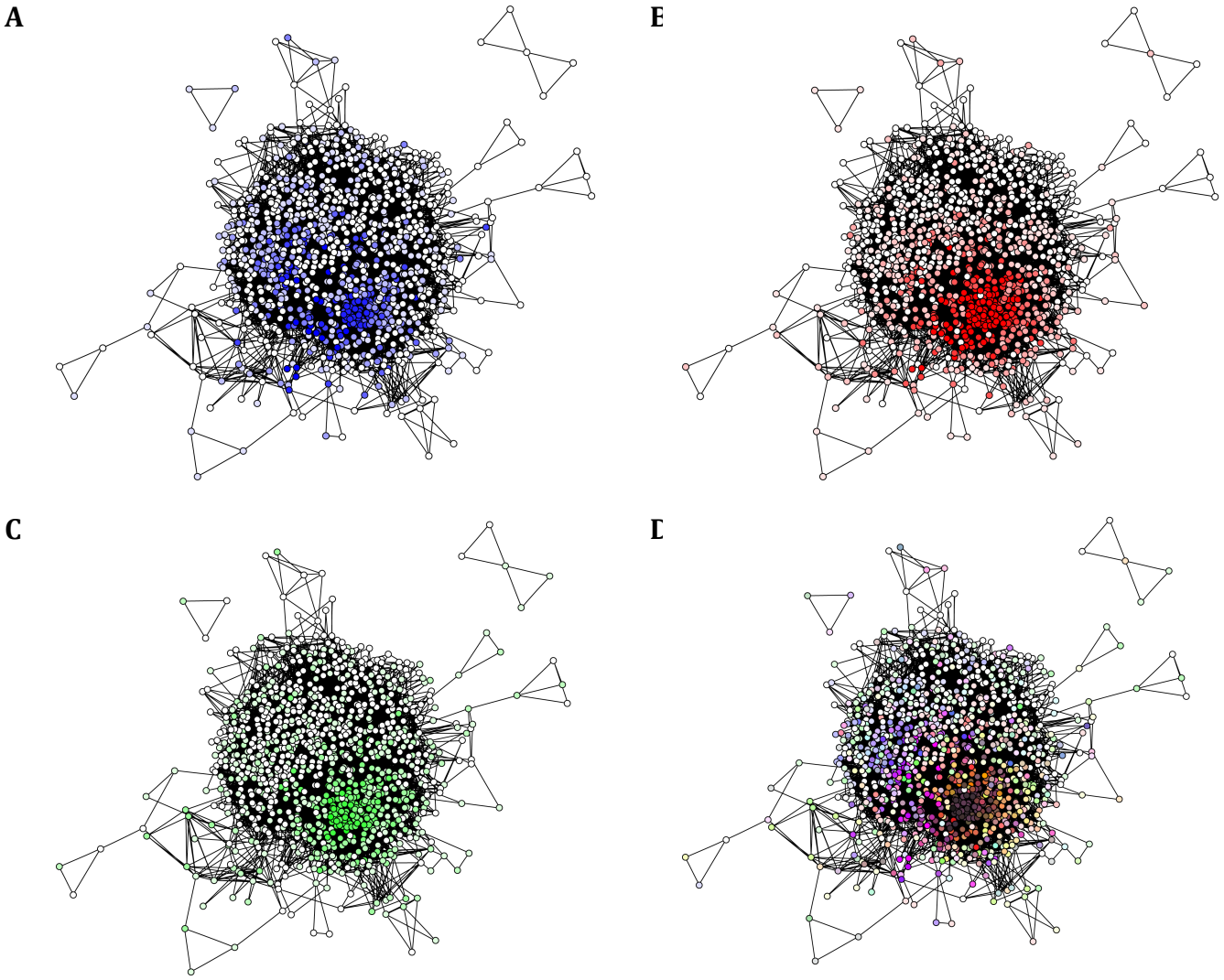


Figure 2. Network view of correlations between signature orderings. Nodes represent each signature tested, and are joined by edges representing the highest positive 1% and negative 1% of median correlations between signature ordering pairs across datasets. Nodes are colored according to the proportion of datasets where they have significant associations with ESR status (A), subtype (B), recurrence (C), and the overlap (D).

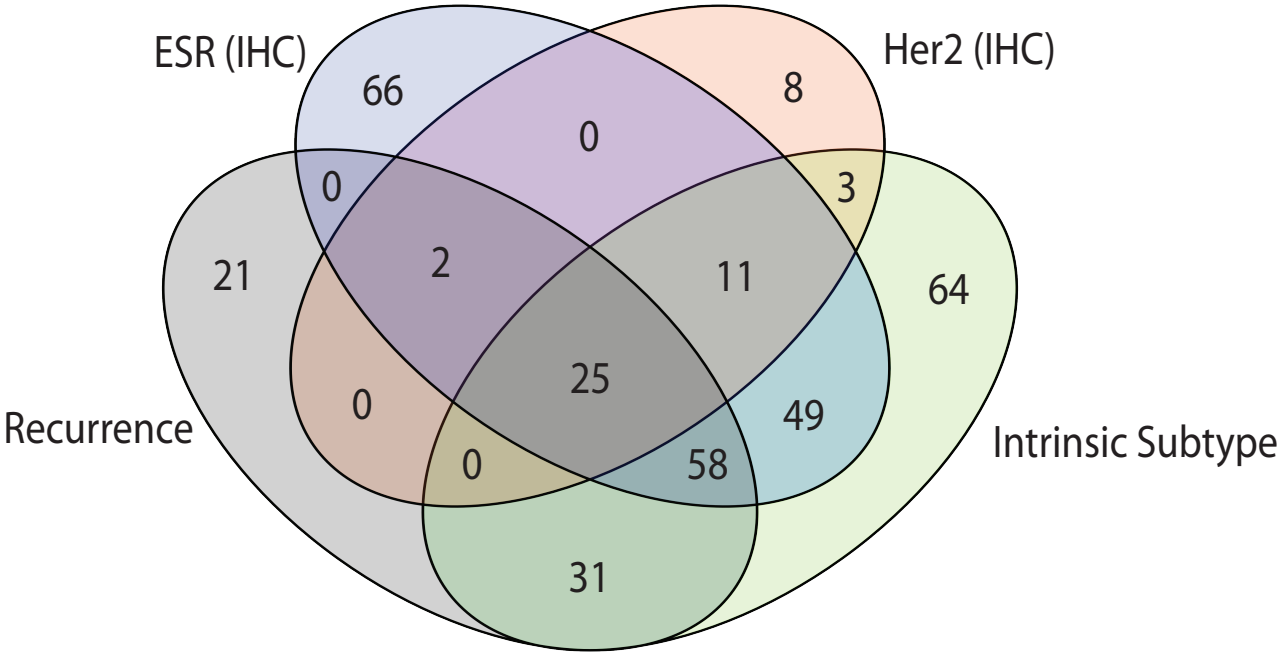


Figure 3. Venn diagram representing significant clinical associations. Signatures must be significantly associated (adjusted p-value<0.05) with ESR status, Her2 status, intrinsic subtype, and/or disease recurrence, in at least 50% of datasets tested. 21 signatures were found to be uniquely associated with recurrence.

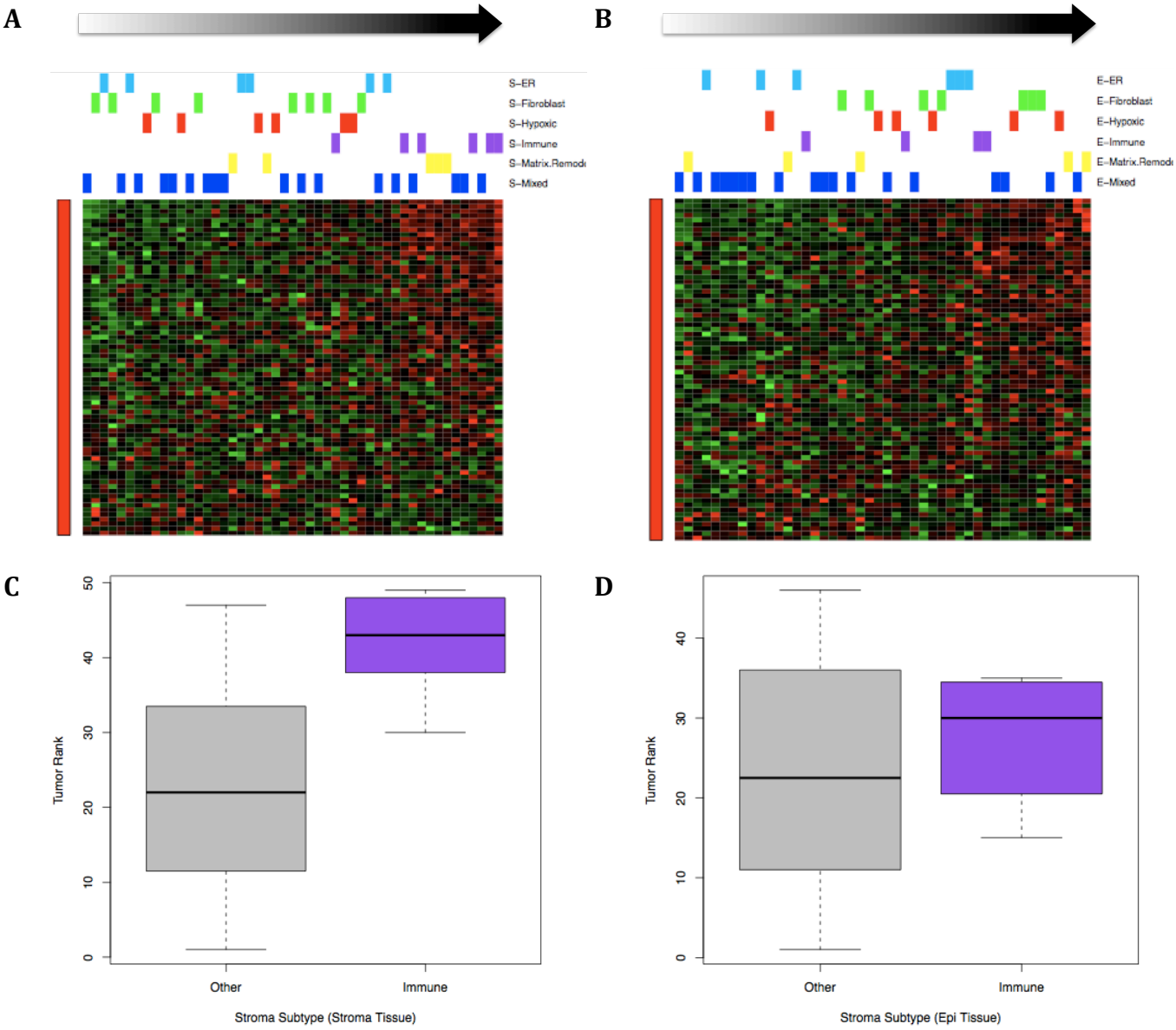


Figure 4. Natural killer cell-mediated cytotoxicity activation signature in stromal and epithelial breast tissue. (A) Heatmap showing laser capture microdissected stromal tissue ordered from left (representing less activation of the signature) to right (representing greater activation of the signature). Samples are labeled according to their intrinsic stromal subtype: ER high (light blue), fibroblast-enriched (green), hypoxic (red), immune-enriched (purple), matrix remodeling (yellow), and mixed (dark blue). (B) Heatmap showing laser capture microdissected epithelial tissue from the same tumors as in A, and labeled according to their intrinsic stromal subtype. Boxplots of the patient rank distributions for immune-enriched (purple) and all other samples (gray) in stromal tissue (C) and epithelial tissue (D). Immune-enriched stromal tissue shows significantly greater activation of the signature (p-vlaue=8.99×10^{-4}), while the epithelial tissue does not (p-value=0.560). Expression data was obtained from [22]; the signature was obtained from [23].

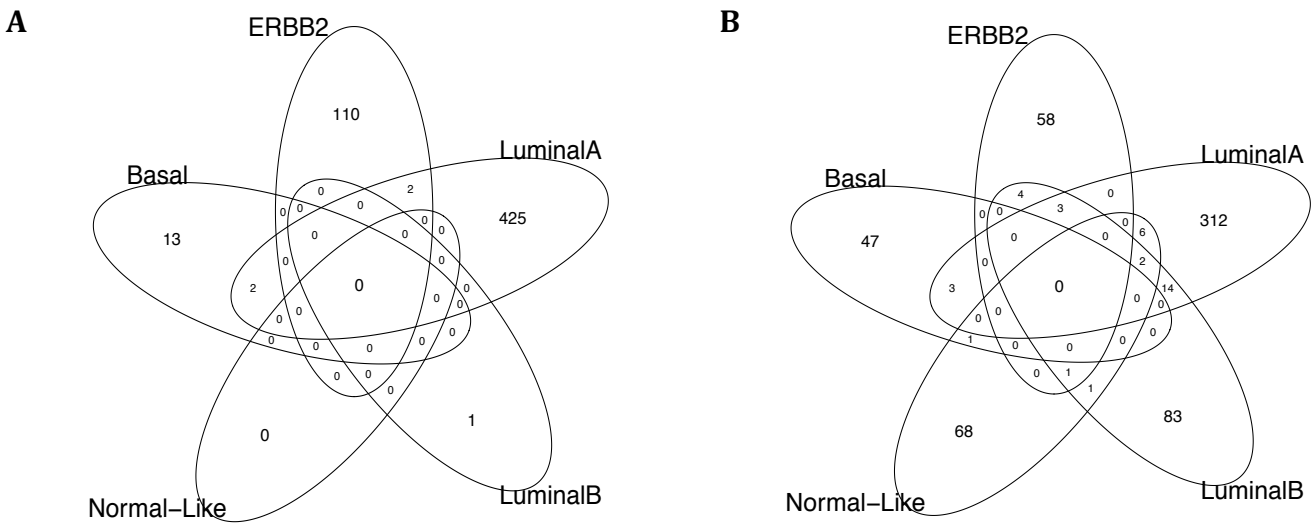


Figure 5. Venn diagram representing the number of univariate genes (A) and multivariate signatures (B) that differentiate significantly DCIS from IDC within each intrinsic subtype, after multiple testing correction. The subtypes demonstrate differences in their number of significant genes and signatures, with very few overlapping between subtypes.

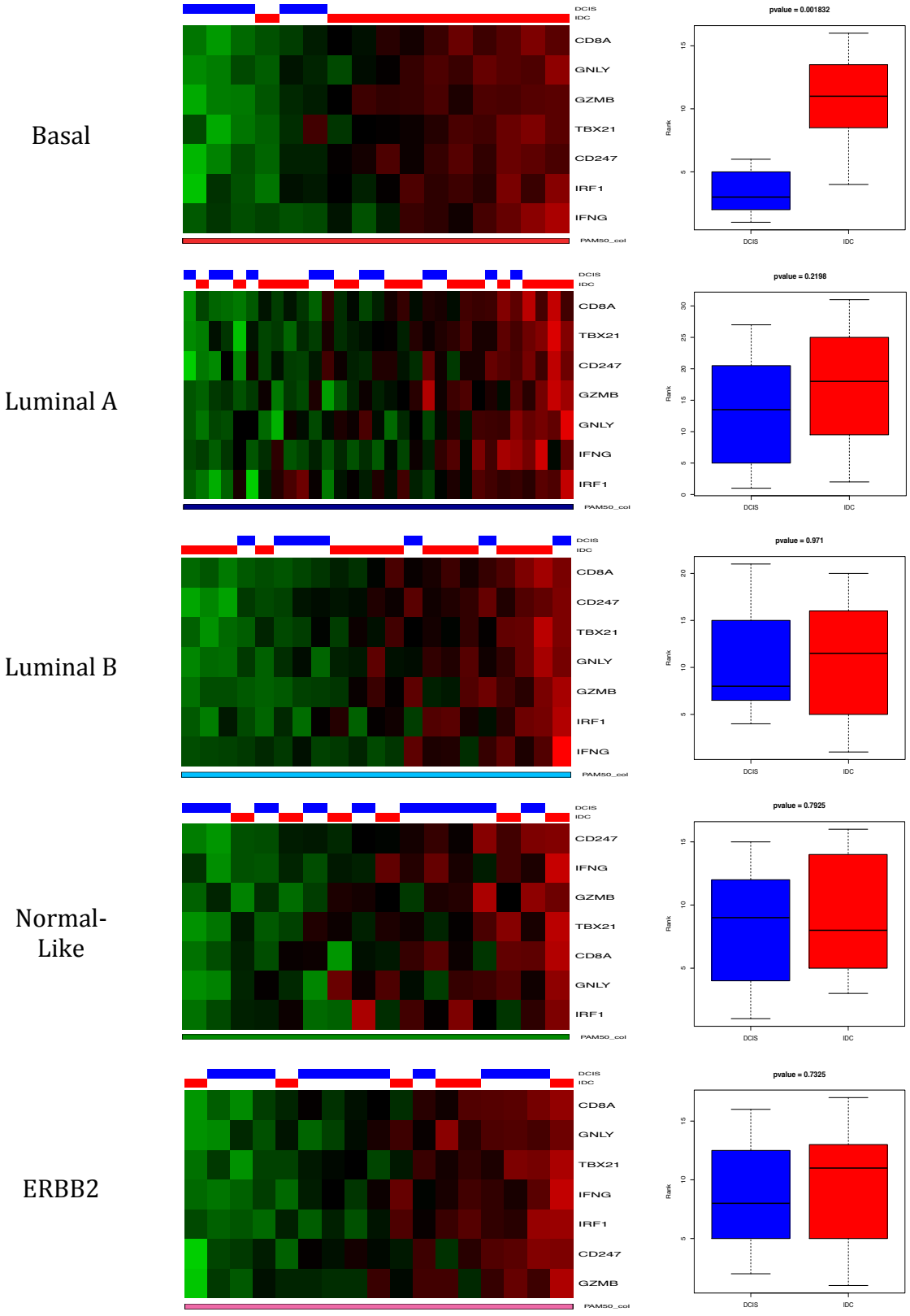


Figure 6. Example of a signature for Th1 adaptive immunity, which specifically differentiates DCIS from IDC tumors in the basal subtype.

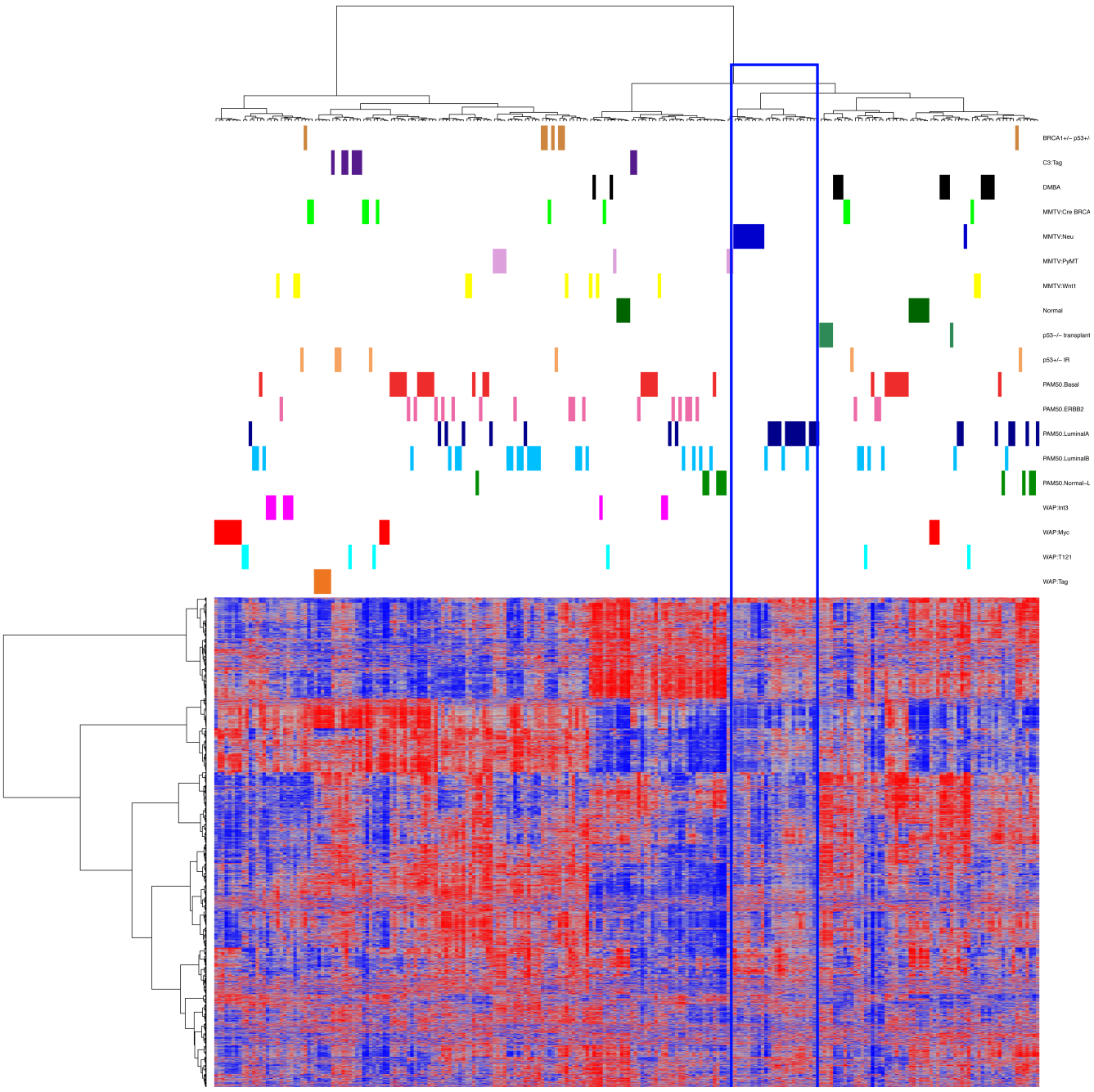
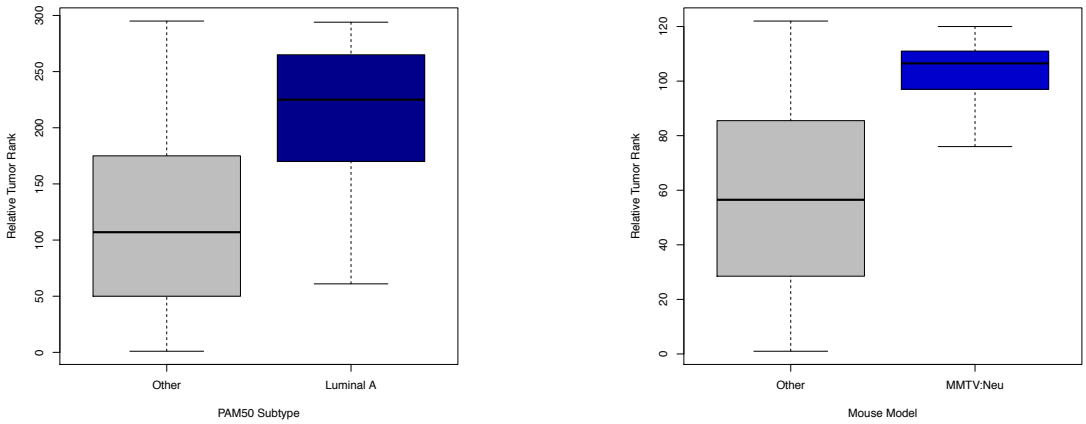
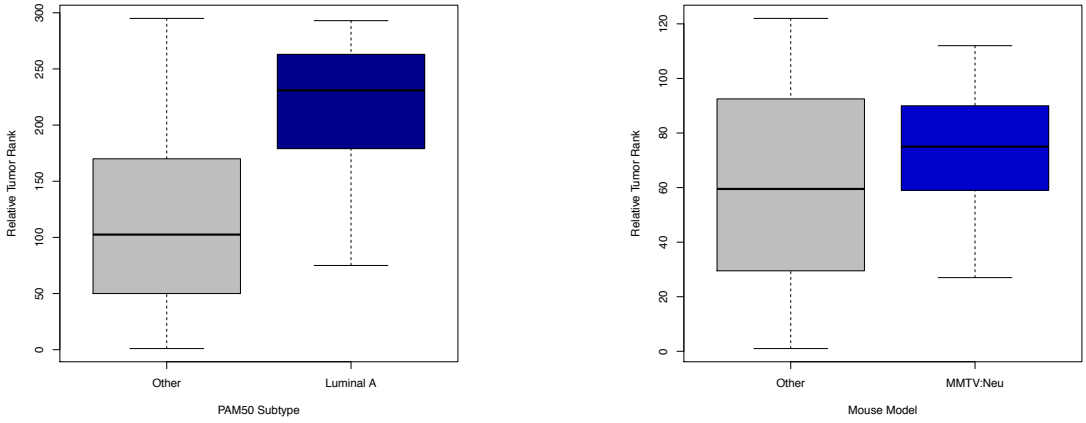


Figure 7. Cross-species hierarchical clustering over ~6400 gene sets. The relative tumor ranks were determined separately for samples in each dataset [24-25], and these ranks are used as features in the rows. The MMTV-Neu mouse model clustered closely with human luminal A tumors (highlighted with blue rectangle). Heatmap is colored from blue to red, representing least to greatest activation of each individual signature respectively.

A



B



C

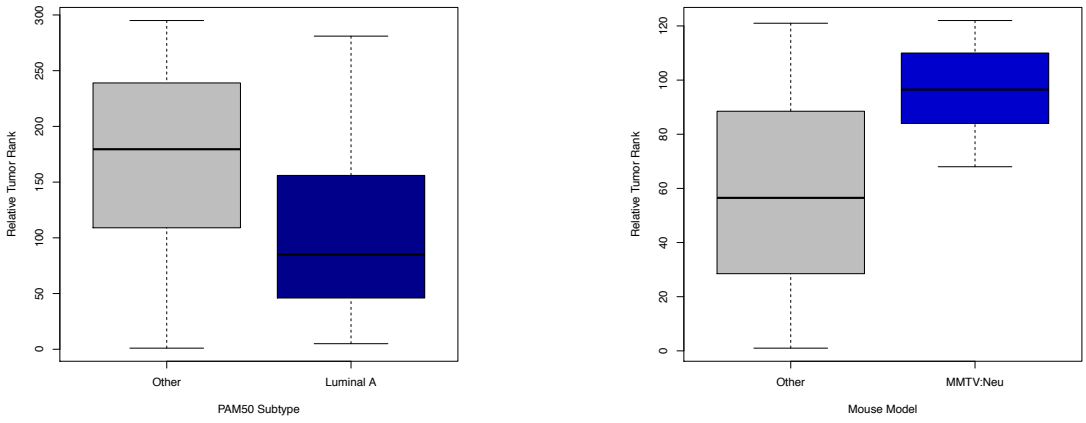


Figure 8. Comparison of relative activation of signatures between human luminal A and murine MMTV-Neu tumors. (A) Both human luminal A and mouse MMTV-Neu display high activation of genes downstream of E2F3. (B) Luminal A tumors display high activation of genes representing response to endocrine signaling, while the mouse tumors do not. (C) MMTV-Neu tumors demonstrate a high transcriptional response associated with interferon activation, while the human tumors do not. Datasets are from [20,25], while signatures were obtained from [26-28].